AROD# 2554:3

NOT RELEASABLE

TO OTS

LECTURES IN APPLIED MATHEMATICS

Proceedings of the Summer Seminar, Boulder, Colorado, 1960

VOLUME I

LECTURES IN STATISTICAL MECHANICS

By G. E. Uhlenbeck and G. W. Ford with E. W. Montroll

VOLUME II

MATHEMATICAL PROBLEMS OF RELATIVISTIC PHYSICS

By I. E. Segal with G. W. Mackey

VOLUME III

PERTURBATION OF SPECTRA IN HILBERT SPACE

By K. O. Friedrichs

VOLUME IV

QUANTUM MECHANICS

By V. Bargmann

LECTURES IN APPLIED MATHEMATICS

Proceedings of the Summer Seminar, Boulder, Colorado, 1960

VOLUME II

by

IRVING E. SEGAL

with

GEORGE W. MACKEY

Mark Kac, *Editor*

The Rockefeller Institute

The Summer Seminar was conducted, and the proceedings prepared in part, by the American Mathematical Society under the following contracts and grants:

Grant NSF-G12432 from the National Science Foundation.

Contract No. AT(30-1)-2482 with the United States Atomic Energy Commission.

Contract Nonr-3081(00) with the Office of Naval Research.

Contract DA-19-020-ORD-5086 with the Office of Ordnance Research.

Library of Congress Catalog Card Number 62-21480

Printed in the United States of America

Foreword

This is the second of a series of four volumes which are to contain the Proceedings of the Summer Seminar on Applied Mathematics, arranged by the American Mathematical Society and held at the University of Colorado for the period July 24 through August 19, 1960. The Seminar was under the sponsorship of the National Science Foundation, Office of Naval Research, Atomic Energy Commission, and the Office of Ordnance Research.

For many years there was an increasing barrier between mathematics and modern physics. The separation of these two fields was regrettable from the point of view of each—physical theories were largely isolated from the newer advances in mathematics, and mathematics itself lacked contact with one of the most stimulating intellectual developments of our times. During recent years, however, mathematicians and physicists have displayed alacrity for mutual exchange. This Seminar was designed to enlarge the much-needed contact which has begun to develop.

The purpose of the Seminar was primarily instructional, with emphasis on basic courses in classical quantum theory, quantum theory of fields and elementary particles, and statistical physics, supplemented by lectures specially planned to complement them. The publication of these volumes is intended to extend the same information presented at the Seminar to a much wider public than was privileged to actually attend, while at the same time serving as a permanent reference for those who did attend.

Following are members of a committee who organized the program of the Seminar:

Kurt O. Friedrichs, Chairman
Mark Kac
Menahem M. Schiffer
George E. Uhlenbeck
Eugene P. Wigner

Local arrangements, including the social and recreational program, were organized by a committee from the University of Colorado, as follows:

Charles A. Hutchinson
Robert W. Ellingwood

The enduring vitality and enthusiasm of the chairmen, and the cooperation of other members of the university staff, made the stay of the participants extremely pleasant; and the four agencies which supplied financial support, as acknowledged on the copyright page, together with the Admissions Committee, consisting of Bernard Friedman, Wilfred Kaplan, and Kurt O. Friedrichs, Chairman, also contributed immeasurably to the successful execution of the plans for the Seminar.

The Seminar opened with an address given by Professor Mark Kac, Department of Mathematics, Cornell University, on the subject "A Mathematician's Look at Physics: What Sets us Apart and What May Bring us Together." Afternoons were purposely kept free to give participants a chance to engage in informal seminars and discussions among themselves and with the distinguished speakers on the program.

Editorial Committee

V. BARGMANN
G. UHLENBECK
M. KAC, CHAIRMAN

Contents

Preface

This book gives the approximate text of a course of eight lectures from combined rigorous mathematical and physically conceptual viewpoints, supplemented by two more purely mathematical lectures. The main purpose is to provide an up-to-date introduction, for the mathematically trained reader, to the central mathematical features of fundamental relativistic physics. While we have aimed for accuracy and scope of perspective rather than for completeness of detail, this purpose itself seemed better served by the inclusion of several detailed discussions and the omission of any significant treatment of many important topics, whose inclusion would not in our judgement have altered the essential form which we have attempted to delineate. In particular, the theory is very largely presented in terms of Bose-Einstein quantum fields, Fermi-Dirac fields being brought in only very briefly and in a descriptive way.

A relatively informal lecture style seemed the best adapted to the quite challenging task of formulating the mathematically intelligible essence of such a complex and sophisticated subject as quantum field and particle theory with the requisite conciseness. No attempt has been made to change this form of presentation in the printed text, in view of its apparent appropriateness for this task.

While the mathematical beauty and inevitability of many parts of modern relativistic physics are now clearly visible, there remain unresolved foundational questions, which in fact dominate the scientific area being considered. It is our conviction that quantum field theory, at least, is on the verge of becoming mathematically firmly established, and will in fact in a few years be recognized as closely parallel to the analytical theory of functionals over infinite-dimensional non-linear manifolds admitting group-invariant differential-geometric structures. In any event, we hope to have given some measure of the recent advances in the subject, and to have conveyed some feeling for the magnificent intuitive scientific structure which has yet to be fully understood mathematically.

Special thanks are due Leonard Gross and David Shale for scientifically useful comments, as well as to the former for help with the original notes.

Varieties of Approaches

To clarify our general intentions and purposes in these chapters, let us review very briefly the varieties of approaches to quantum fields and particles which are currently popular. Although the *ultimate* aims of many theoreticians are rather similar, involving mainly an increase in our understanding of fundamental physical phenomena, their shorter-term objectives are quite varied, so much so that fundamental theoretical physics has a rather fragmented appearance at present.

The traditional approach effectively regarded theoretical physics as a game whose purpose was to derive from simple theoretical principles the abstruse numbers obtained in laboratory experiments on particles; this description is a variant of one due to Dirac. The great success of Dirac, Heisenberg, Schrödinger, and many others at this game during the late twenties laid incidentally the foundations of modern quantum theory. But in the past thirty years the game has proved so difficult that people have generally felt forced to modify its rules in one way or another.

The success of the renormalization theory initiated in clear-cut form chiefly by Feynman, Schwinger, and Tomonaga, in computing with great accuracy quantum radiation effects on the electron, represents the most remarkable theoretical explication of fundamental physical data in the past thirty years. It was based however on a certain relaxation of the rules permitting the use of an ad hoc argument at a crucial stage in the computation to resolve a serious difficulty, i.e. eliminate the so-called divergences to which the theory and mathematical procedure led. This remains the case today despite the considerable simplifications and clarifications due to Dyson, Ward, Salam, van Hove and his associates, and many others.

More recently the "axiomatic" schools which have emerged from this situation have surpassed the traditional approach in logical clarity, utilizing an explicit rather than implicit statement of their fundamental principles. They have concentrated on increasing understanding of the meaning, scope, and general implications of quantum field theory, and have effectively given up the attempt to compute experimental data from theoretical principles. The most active of these "schools," including notably that of Källén-Wightman and that of Lehmann-Symanzik-Zimmermann (to both of which Haag and Jost have made significant contributions), are rather mathematical in spirit but do not always distinguish between mathematically rigorous and partially heuristic definitions and results. From an overall point of view, however, the main problem here is the lack as yet of non-trivial examples of systems satisfying the axioms, i.e. systems involving real emission and absorption of particles.

Roughly at the other end of the theoretical spectrum from the axiomatic schools are those concerned chiefly with the correlation of experimental data by means of approximations to and heuristic techniques in quantum field theory of varying degrees of physical motivation and, unfortunately, quite uncertain reliability. In any event, the ideas of Chew, Goldberger, and Low have proved to be particularly useful in reducing the large and rapidly growing volume of experimental data in nuclear physics. The technique of so-called "dispersion relations" has been widely used for a substantial time, and some of the relations have been supported by experimental evidence, but a clear-cut formulation and derivation of the relations within a rigorous mathematical framework has not yet been given, and it also seems quite difficult in the nature of things to make a conclusive experimental test of the relations, since, unlike the familiar relations that have been so tested, the checking of an individual numerical equality in a dispersion relation necessarily involves measurements at all, including arbitrarily high, energies.

These three schools have certain connections, a particularly interesting and actively investigated one being that between the empirically-oriented and the axiomatic schools via the theory of dispersion relations. But on the whole there does not appear to be much prospect for their fundamental unification in the foreseeable future. On the other hand, until the elementary question, of what, precisely, a quantum field theory consists of, is answered in satisfactory physical and mathematical terms, there are insufficient rational grounds for pessimism or optimism.

A pure mathematician who is interested in fundamental physics will see at once that there is another possible approach, that of building up on the bedrock of rigorous mathematics, while keeping as close as possible to the ideas that emerge from empirical practice. Ten years ago such an approach might have seemed very naïve, but by now it is clear that the rigor and mathematical method, far from proving burdensome, enable one to deal simply and definitely, if in a rather sophisticated way, with some of the really significant theoretical questions; and that the close connection between that which is mathematically viable and physically meaningful is a rather general feature of the situation, and not limited to such cases as that treated by Bohr and Rosenfeld in their classical work on the measurability of the electromagnetic field.

Our purpose here is mainly to treat those parts of the theory of fields and particles which are now available in a rigorous, compact, and general form. The solution of the relevant problems has tended to lead to new problems, some of which we shall describe. We shall have to pay the price of increasing at least temporarily the difficulty of making a dictionary for translating between experimental physics and mathematics. We must not expect too much direct physical contact too soon, in view of the very substantial complications inherent in any comprehensive theory conceivably applicable to elementary particle interactions. But the pursuit of this game of capturing modern physical ideas and principles in rigorous and simple mathematics is a reasonable and interesting activity in itself. We think moreover that there are now visible lines of development offering definite promise of dealing effectively with physically interesting relativistic interactions.

From a purely mathematical point of view the main mathematical fields pertinent to the general theory of particles and fields are:

1. Operator theory (especially operator algebras).
2. Theory of group representations (especially of the Lorentz and other physical symmetry groups).
3. Theory of functionals.
4. Theory of partial differential equations.

Large parts of these subjects are relevant here, in fact a year's course on each of them would not be amiss. Of course, here we can treat only a few aspects of special relevance. We shall say only a little about operator theory, and less about group representations, as these will be treated in Professor Mackey's chapters. We shall discuss analysis in function space, because of its relevance and relative novelty, and note its relation to the line of development originating with the work of Wiener on Brownian motion. We shall do little with the theory of partial differential equations, partly because the aspects of the theory of greatest relevance—the global spectral theory of variable coefficient and non-linear hyperbolic equations—are as yet rather undeveloped.

CHAPTER I

Quantum Phenomenology

We begin by treating the notion of a physical system, keeping as closely as possible to the use of concepts having a fairly direct empirical or physically intuitive significance. The fundamental object associated with a physical system may be taken either as an *observable* or as a *state*. The former concept seems simpler from a naïve point of view, and leads to a viable theory in terms of which state may be treated quite effectively, so we shall start with observable as a fundamental undefined notion.

We should mention parenthetically that the early formulation of quantum phenomenology asserted that: (1) an observable is a self-adjoint operator in a Hilbert space; (2) a state is a vector ψ in this space; the connection between (1) and (2) being that the expectation value of A in the state ψ is $(A\psi, \psi)$. These "axioms" are technically simple, but they are thoroughly unintuitive and ad hoc. In addition, it has turned out recently that they are technically really effective only in the case of systems of a finite number of degrees of freedom. In fact certain of the ultraviolet divergences of quantum field theory result indirectly from the inadequacy of the older phenomenology. Therefore there is ample reason, both foundational and technical, to prefer the more recent form, which is given below.

Now both physically and mathematically it appears that the bounded observables play the fundamental role, the unbounded ones being readily dealt with in terms of the bounded ones, as far as foundational purposes are concerned. Taking e.g. a one-dimensional quantum-mechanical particle, no given finite physical apparatus can conceivably accurately measure the momentum p, once this momentum goes beyond a certain limit. Now one may construct larger and more refined apparatus, and thereby for each finite n, measure $F_n(p)$, where $F_n(x) = x$ for $|x| \leqq n$ and $F_n(x) =$ say $n \operatorname{sgn} n$ for $|x| > n$. That is, one can measure the infinite sequence of observables $F_1(p), F_2(p), \cdots$, each of which is bounded; and p itself is not measurable directly but only as a limit of such a sequence, and so involves an unphysical infinity of experiments. On the other hand, mathematically we are entitled to look ahead a bit, to the first axioms of early quantum phenomenology, to the effect that an observable is a self-adjoint operator in a Hilbert space. The well-known great difficulty of performing effectively simple algebraic operations on non-commuting self-adjoint operators together with the various possibilities for treating unbounded in terms of bounded operators, strongly suggest the limitation to the bounded ones.

So we consider to begin with only the bounded observables of the given physical system. From an intuitive point of view it is clear that if A is a bounded observable and α is a real number, then αA is a bounded observable; it is measured simply by measuring A and multiplying the result by α. Similarly A^2 is a bounded observable, measured by measuring A and squaring the result. Now if B is another observable, the sum $A + B$ and product AB can be similarly defined only when A and B are simultaneously observable. We may however define $A + B$ in a more indirect physical fashion as that observable whose expectation in any state is the sum of the expectations of A and of B. Intuitively it is plausible that an observable may be reconstructed from its expectation values in all states; alternatively this definition may be regarded as a restriction on the states of the system. On the other hand, the product AB may not be defined in a similar fashion because it is not even true for simultaneously observable A and B that the expectation value of the product is the product of the expectation values (as is familiar in the theory of probability, whose observables are usually called "random variables").

Thus it is physically reasonable to postulate that the bounded observables of the physical system form a type of algebra, the relevant operations being multiplication by scalars, squaring and addition of observables, but not multiplication in general. However, in view of the indirect character of the definition of addition, the full reasonableness of the assumption that two observables can be added will follow only if the theory which is built up from such assumptions has as a logical consequence the rationalizing assumptions that the expectation value of the sum of two observables is the sum of their expectation values, in a particular state, and that any observable can be recovered from its expectation values.

In addition it is reasonable to assume that there is a unit observable I whose expectation value in every state is unity, and that the usual rules for the reduction of measurements of simultaneously measurable observables are valid. This last requirement turns out to be needed only in the form

$$A^r \circ A^s = A^{r+s}, \qquad (\alpha A)^r = \alpha^r A^r$$

if the pseudo-product $A \circ B$ is defined by the equation

$$A \circ B = \frac{1}{4}[(A + B)^2 - (A - B)^2],$$

and A^r is defined recursively by the equations

$$A^0 = I, \qquad A^r = A \circ A^{r-1}.$$

Since $A \circ B$ coincides with the phenomenological product described above when A and B are simultaneously observable, the present requirements have immediate intuitive validations.

Thus we have rationalized the following mathematical axiom:

PHENOMENOLOGICAL POSTULATE, ALGEBRAIC PART: *A physical system is a collection of objects, called (bounded) observables, for which operations of multiplication by a real number, squaring, and addition are defined, and satisfy the usual assumptions for a linear vector space as well as those involving the squaring operation given above.*

As a mathematical example, consider the set of all bounded hermitian (linear, everywhere defined) operators on a Hilbert space. It is obvious that with the usual algebraic operations the foregoing postulate is satisfied. It may be helpful to note incidentally that the conventional product of operators is not meaningful within this system, since the product of two hermitian operators will again be hermitian only when they commute; while the pseudo-product $A \circ B = (AB + BA)/2$ in the present case has for non-commuting hermitian A and B no physical interpretation.

Now the main result we need in the present connection is the

BASIC PHENOMENOLOGICAL PRINCIPLE: *Any physical system is determined in all its physically observable aspects by its algebra of bounded observables.*

That is to say, two systems whose bounded observables may be brought into one-to-one correspondence, in such a fashion that sums, squares, and products by real numbers correspond, are physically identical—apart from the labelling of the observables.

To explain more precisely what is meant by a "physically observable aspect," let us introduce the key notions of state, pure state, and spectral (exact possible) value of an observable. From an empirical standpoint, a state E exists only as a rule which assigns to each bounded observable its expectation value in the state; any possible metaphysical distinction between the state and the corresponding functional on the observables is irrelevant for empirical objectives. Accordingly, we define a state as this functional and the following properties of a state E have a clear intuitional validity:

1. *Linearity*: $E(A + B) = E(A) + E(B)$,
$E(\alpha A) = \alpha\, E(A)$,

if A and B are any bounded observables and α is a real number.

2. *Positivity*: $E(A^2) \geqq 0$.
3. *Normalization*: $E(I) = 1$.

Thus on a rather conservative physical basis a state must be some sort of normalized positive linear functional on the observables. For basic phenomenological purposes this is all that turns out to be required for a state, and so we follow von Neumann in *defining* a state as such a functional.

Now if a system is in a state E with probability α and in a state E' with probability α', where $\alpha + \alpha' = 1$ and $\alpha > 0$, $\alpha' > 0$, the effective state of the system is E'', where

$$E''(A) = \alpha E(A) + \alpha' E'(A).$$

The state E'' is called a *mixture* of the states E and E' and following Weyl we call a state *pure* if it cannot be represented as a mixture of two distinct states. It is evident that it is the pure state that plays the fundamental part in non-statistical mechanics; an experiment of maximal theoretical accuracy will yield a pure state of the system.

To clarify these notions, consider briefly the system of all bounded hermitian operators in a Hilbert space $\mathscr{H}$. If ψ is any unit vector in $\mathscr{H}$, the functional E defined by the equation

$$E(A) = (A\psi, \psi)$$

is easily seen to be a state. It is actually a pure state, as can be seen in a fashion that will be indicated later. In many conventional treatments of quantum mechanics, the vector ψ is called a state, but it is evidently of quite another character from the functional E which is here defined as a state. In particular ψ is incompletely physically observable, any multiple of ψ by a number of unit modulus being physically indistinguishable from it. Here ψ will be referred to as a *state vector* or *wave function* for the state E. Incidentally, it is only in the trivial case of the finite-dimensional Hilbert space $\mathscr{H}$ that every pure state has the foregoing form; for an infinite-dimensional space there are others, which can arise, e.g., from the continuous spectrum which may manifest itself for operators in an infinite-dimensional space.

An example of a mixed state is provided by one of the form

$$E(A) = \operatorname{tr}(AD),$$

where D is a non-negative self-adjoint operator of absolutely convergent trace, and total trace unity. D is then uniquely determined by E and is called the "von Neumann density operator (matrix)." Such a state is pure if and only if D is of unit rank, in which case it arises from a wave function in the fashion just indicated.

To arrive at the notion of spectral value, the variance of an observable A in a state E may be reasonably defined as the quantity $E(A^2) - E(A)^2$, which is automatically non-negative by virtue of the positivity of the functional E. In line with this, A may be said to *have an exact value in the state* E in case its variance vanishes; and the values $E(A)$ of A in all such states designated as the spectrum of the observable A.

Now it is clear from the definitions of state, pure state, and spectral value, that they are wholly determined by the algebra of bounded observables. But this is significant only if states and pure states exist in ample number, and if spectral values exist, and relate to states in the usual probabilistic fashion (i.e. the expectation of the observables is the average of the spectral values with respect to a probability distribution determined by the state), etc. To prove such results the phenomenological postulate above must be supplemented by a postulate making possible the application of analytical methods.

To arrive at a physically meaningful postulate that will be mathematically effective, consider the properties which may be anticipated for the bounds of the observables, whose finiteness has not thus far been utilized. The bound represents, in an intuitive physical way, the greatest possible absolute value for the observable. This interpretation together with a quite moderate amount of reflection shows the physical basis for the

PHENOMENOLOGICAL POSTULATE, ANALYTICAL PART: *To each observable A is assigned a "bound," designated $\|A\|$, in such a way that the following conditions are satisfied:*

i. $\|A\| \geqq 0$ *and* $\|A\| = 0$ *if and only if* $A = 0$.

ii. $\|\alpha A\| = |\alpha|\,\|A\|$ *and* $\|A + B\| \leqq \|A\| + \|B\|$.

iii. *The collection of all observables is complete with respect to the metric determined by the bound, i.e. if $A_1, A_2, \cdots$ is a sequence of observables such that $\|A_m - A_n\| \to 0$ as $m, n \to \infty$, then there exists an observable A such that $\|A_n - A\| \to 0$.*

iv. $\|A^2\| = \|A\|^2$ *and* $\|A^2 - B^2\| \leq \text{Max}\,[\|A^2\|, \|B^2\|]$.
v. A^2 *is a continuous function of* A, *i.e. if* $A_n \to A$, *then* $A_n^2 \to A^2$.

Conditions i and ii have a direct physical justification. The condition iii is virtually a matter of convenience, for an incomplete system could always be completed; an observable A could be defined if necessary, its expectation value in a state being explicitly obtainable as the limits of the expectation values of the A_n. Condition iv takes a slight amount of reflection for its intuitive justification. Condition v merely asserts that if two observables are close (as measured by the bound of their difference) then so are their squares.

For an example, consider again the system of all bounded hermitian operators on a Hilbert space, with $\|A\|$ defined as the usual bound of the operator A. That is, $\|A\|$ is the least upper bound of the Hilbert space norms $\|A\psi\|$ as ψ varies over all unit vectors, or equivalently, for hermitian operators, of $|(A\psi, \psi)|$. All of the foregoing conditions follow almost trivially.

On the strength of the combined algebraic and analytical parts of the phenomenological postulates, all of the physically plausible and conventionally accepted principles of quantum phenomenology may be rigorously established. The proofs are based on now familiar results and methods of abstract analysis, including notably the Stone-Gelfand representation theory and such results in linear analysis as the Hahn-Banach, Krein-Milman, and Riesz-Markoff theorems.

Among the results are:

1. *There exists an ample supply of pure states, in the sense that two observables having the same expectation values in all pure states must be identical.* In particular, the justification for the assumption that two observables can be added is completed.

2. *Any observable admits a closed set of spectral values, and the expectation of the observable in any state is the average of these spectral values with respect to a probability distribution on them canonically determined by the state.* Specifically, this distribution may be defined as that with characteristic function $E(e^{itA})$, where E is the state and A the observable (here e^{itA} is defined in the obvious fashion, or alternatively, $E(e^{itA})$ may be replaced by $E(\cos tA) + iE(\sin tA)$, where $\cos tA$ and $\sin tA$ are defined by the demonstrably convergent, conventional power series expansions). It is not difficult to see that this function (of t) is positive definite and the Fourier-Stieltjes transform of a probability distribution.

3. *The smallest closed system of observables* (*in the sense of the*

***phenomenological postulates) containing a given observable A is in* 1–1 *algebraic correspondence with the algebrà of all continuous functions on the spectrum of A*.** This gives a spectral representation for observables quite analogous to that for self-adjoint operators on a Hilbert space, in fact the latter may be in part deduced from the former. From this together with the Riesz-Markoff representation theorem for positive linear functionals on spaces of continuous functions a definition for the probability distribution on the spectral values of an observable, in a given state, follows directly without consideration of the characteristic function.

In the case of the system of all bounded hermitian operators and a pure state arising from a wave function ψ, in the fashion indicated above, the conventional definition for the probability distribution which assigns to the operator with spectral resolution $\int \lambda dE_\lambda$ the probability distribution with element $d\|E_\lambda\psi\|^2$ can readily be shown to be in agreement with the present one, which depends only on the algebra of the operators, and not at all on the specific fashion in which the operators are represented on a Hilbert space.

There is one further result of methodological importance. In mathematical and physical practice it is convenient, and possibly inevitable in the nature of things, to restrict the scope of the physical system under consideration. It is conceivable that when this limited physical system is replaced by a larger system, e.g. the "universe," the physical existence of states and/or the spectral values of observables might be affected. There might for example be a selection rule prohibiting certain pure states of the convenient subsystem because they could not be realized in the larger system from which the subsystem could not really be isolated. This would make the situation a very complicated one, and it is therefore good to know that it is mathematically demonstrable that this difficulty cannot arise:

4. *Any pure state of a physical system which is a subsystem of a larger system can be realized in a pure state of the larger system.* (That is, there exists a pure state of the larger system which coincides on the subsystem with the given pure state.) In particular, the spectral values of an observable are independent of the algebra of observables of which it is considered to be a member—as is essentially true also of probability distributions for the spectral values.

5. *The bound of an observable A may be defined purely algebraically as the least real number α* such that $\alpha I - A = B^2$ and $\alpha I + A = C^2$ for suitable observables B and C. Intuitively, this defines $\|A\|$ as the least α such that $\alpha I \pm A$ is non-negative, as is evidently intuitively justifiable.

(We could in fact have given a formally algebraic set of axioms by introducing the bound in this way, etc., but there is no real advantage in this.) Thereby we complete the proof of the basic phenomenological principle stated earlier, that all purely phenomenological features of a physical system are determined by the algebra of the bounded observables.

The question now arises as to the character of the mathematical systems satisfying the phenomenological postulates. It might be extremely difficult to classify all such systems, and not necessarily rewarding, for the exclusive relevant prototype is the kind of system in which the pseudo-product $A \circ B$ is bilinear in A and B, or more specially, which consists of all self-adjoint elements $C = C^*$ of an associative algebra on which there is defined an adjunction operation $C \to C^*$ satisfying the usual type of requirements ($C^{**} = C$, $(\alpha C)^* = \bar{\alpha} C^*$, C^*C positive in a certain sense for $C \neq 0$). These assumptions have no quantitative empirical justification whatsoever, but they are simple and natural from a purely mathematical viewpoint, and it is interesting that they force the observables to have interpretations as hermitian operators in a Hilbert space, there being in general, however, nothing unique about the interpretation or the Hilbert space. In the case where it is assumed only that the pseudo-product is bilinear the situation has not yet been fully analyzed but Sherman has shown that the exceptional simple Jordan algebra of Albert satisfies the postulates. There is at present no evidence of any physical relevance for this system, whose finite-dimensionality also sharply limits its conceptual interest.

All definite theoretical systems of observables that have thus far been proposed are in fact representable in terms of operators on a real or complex Hilbert space, and we may as well restrict ourselves at this point to such systems, of which there are many. To be more specific, we define as a concrete C^*-algebra $\mathscr{A}$, an algebra of bounded linear operators on a real or complex Hilbert space, which is closed under the adjunction operation, and also in the uniform topology (i.e. contains all limits of uniformly convergent sequences of operators in $\mathscr{A}$, where A_n converges to A uniformly in case $\|A_n - A\| \to 0$, the operator bound being defined as above). Now two concrete C^*-algebras may be algebraically isomorphic (in one-to-one correspondence in a fashion making sums, products, and adjoints correspond) without there being any simple connection whatsoever between the Hilbert spaces on which the respective operators act. The relevant object here is an *abstract* C^*-algebra, which may be defined as an equivalence class of C^*-algebras under algebraic isomorphism. The set of all self-adjoint elements of an abstract C^*-algebra forms then a physical system, as defined above, the bound being defined in the algebraic manner indicated above.

The complete description of a physical system involves however not only the statement of the mathematical character of the algebra of bounded observables, but also a labelling of the observables, a kind of physical-mathematical dictionary. This is clearly visible e.g. in the fact that in elementary quantum mechanics it is assumed that the bounded observables consist of all bounded hermitian operators on a countably-dimensional Hilbert space, irrespective of the number of degrees of freedom of the system.

Now there is evidently no mathematical labelling scheme that will be applicable to a perfectly general C^*-algebra of observables. However the physically relevant C^*-algebras all involve implicitly or explicitly a labelling scheme whose mathematical structure is of essential importance in the theory. In its simplest form this labelling involves the designation of certain observables as "canonical," or as "field variables." The treatment of these labelling matters involves additional elements of mathematical structure of quite a different character from those of the present lecture, and will be gone into in the next lecture. We can however give a rough indication of what is involved as well as illustrate pure phenomenology by considering briefly the Heisenberg commutation relations in one dimension.

In their original form

$$pq - qp = \frac{h}{i},$$

the commutation rule left open considerable room for irrelevant mathematical pathology and it is clear by now that it is the Weyl form of the rule, where units are used for which $h = 1$,

$$e^{i\alpha p}\, e^{i\beta q} = e^{i\alpha\beta}\, e^{i\beta q}\, e^{i\alpha p}$$

(α and β being arbitrary real numbers), that is relevant in a rigorous approach. A canonical pair is, then, an ordered pair (p, q) of self-adjoint operators in a complex Hilbert space which satisfies the cited (Weyl) relations. The Schrödinger representation

$$e^{i\alpha p}: f(x) \to f(x + \alpha), \qquad e^{i\alpha q}: f(x) \to e^{i\alpha x} f(x)$$

(f an arbitrary square-integrable function on $(-\infty, \infty)$) in the Hilbert space $\mathscr{H} = L_2(-\infty, \infty)$, shows that canonical pairs exist. Here we use Stone's theorem that a continuous one-parameter unitary group, such as the group

$$U_\alpha : f(x) \to f(x + \alpha),$$

has a self-adjoint generator; the continuity here is in the weak topology, which is equivalent to the assertion that $(U_\alpha f, g)$ is a continuous function of α for all f and g in $\mathscr{H}$.

Consider now the question of what should be the bounded observables associated with a canonical pair. Conventionally one regards all bounded hermitian operators as such, in the case of the Schrödinger representation; in general in a situation of this sort, people commonly form the "ring of operators" in the sense of Murray and von Neumann that is generated by the basic operators, and regard the self-adjoint members of this ring as observables. This ring of operators consists of all bounded "functions" of the basic operators, a function of a set of operators being defined as an operator that commutes with every unitary operator U as well as with U^* for all U commuting with every operator in the given set. The Schrödinger representation is irreducible (i.e. there is no non-trivial closed linear subspace of $L_2(-\infty, \infty)$ that is invariant under all the $e^{i\alpha p}$ and $e^{i\beta q}$), as follows from the "ergodicity" of the action of the group of all translations on the reals, i.e. the absence of any non-trivial measurable sets invariant within sets of measure zero under all translations. Now it follows from the von Neumann commutor theorem about rings of operators that no unitary operator can commute with both p and q in the Schrödinger representation other than the trivial operators αI, where α is a constant of unit modulus. It follows in turn that every bounded operator on the representation space is a function of p and q.

Alternatively, one may define "ring of operators" as a collection of bounded operators, including the identity operator I, which is closed under the usual algebraic operations of addition, multiplication, and adjunction, and also closed in the *weak topology*. The latter is defined by designating an operator as a limit point of a set of operators in case any finite set of matrix elements of the operator can be matched within an arbitrary ϵ by the matrix elements of some operator in the set. The von Neumann commutor theorem, one version of which is to the effect that any operator commuting with all unitary operators commuting with every operator in a given ring of operators, is itself in the ring, shows that the ring of operators generated by a set of bounded operators as defined above is the same as the smallest ring of operators containing the set. Thereby the ring of operators generated by the canonical pair may also be defined as the set of all limits in the weak topology of the ring of all finite linear combinations of the

$$e^{i\alpha p}\, e^{i\beta q}.$$

The point of this technical digression is partly informational, and partly to indicate how remote the operators in the weakly closed ring generated by a set of operators may be from any apparent empirical connection with the basic set. This somewhat unphysical character for the weakly closed ring does not cause any harm in the case of systems of a finite number of degrees of freedom, but it does cause difficulty in the case of infinite systems, the discrepancy being due to, or, more precisely, having a counterpart in, the validity of the Stone-von Neumann Theorem on the uniqueness of the Schrödinger operators in the one case and its failure in the other. We shall discuss this more fully in the next chapter. It may nevertheless be useful to sketch here briefly how one could deal with the present case in a physically sounder fashion.

If $f(\alpha, \beta)$ is an integrable function of the real variables α and β, then

$$T = \int_{-\infty}^{\infty}\!\!\int \exp\,[i(\alpha p + \beta q)]f(\alpha, \beta)\, d\alpha\, d\beta$$

will be a well-defined bounded operator. (Here we use the definition, $f(A) = f(\bar{A})$, if A is an operator whose closure $\bar{A}$ is normal, and f is a Baire function, to avoid possible ambiguity due to the lack of closure of $\alpha p + \beta q$.) It is clear that T is constructed from p and q in a relatively explicit manner and so may be designated as an explicit function of p and q. The collection of all such T, together with their *uniform* limits (uniform convergence of observables having direct physical meaning) form a more conservative algebra of observables for a one-dimensional quantum-mechanical system than the much larger weakly closed ring described above.

There is an alternative definition for the collection of observables just designated which serves to reinforce its propriety. It may be argued, from an intuitional physical standpoint, that since large values of p and q are measurable only with increasing difficulty, the real observables are not the p or q or exponentials, which depend upon all values of p or q, but "cut-off" functions $f(p)$ or $f(q)$, where f is a continuous function that is constant outside of a finite interval. For greater symmetry and smoothness we may deal with the products $f(p)g(q)$, where f and g are both of the type described; the smallest

C^*-algebra containing all of these can be shown to be the same as that defined in the preceding paragraph (in particular, if p and q are interchanged the same algebra is obtained). It is convenient and creates no difficulty to add I to the algebra, obtaining an algebra whose self-adjoint elements satisfy the phenomenological postulates.

The elements of this algebra of, say, *smooth* observables, are evidently labelled in a relatively clear-cut way in terms of p and q. As a pure algebra it consists of all completely continuous operators on the Schrödinger representation space, apart from additive multiples of I. It may be shown that the states of this algebra (excluding the multiples of I, which are inessential, for simplicity) are *regular* in having the form

$$E(A) = \operatorname{tr}(AD),$$

where D is a hermitian operator of absolutely convergent trace (i.e. has purely discrete spectrum, with an absolutely convergent series of proper values, taking account of multiplicities). It is reassuring that these are precisely the states that are generally considered to be physically truly realizable. Of course, for an arbitrary bounded hermitian operator A, $\operatorname{tr}(AD)$ will exist, and may be called the expectation value of A in the state E, but this is made possible by the felicity of the mathematics, rather than indicated by any physical considerations. In particular, while the conventional practice of regarding all bounded hermitian operators as observable is thereby justified in part, it is important to recognize that not all states of the larger system will be truly physically realizable, the proper state space being the same as that of the smaller system, and consisting only of those of the simple analytical form indicated.

Notes to Chapter I

1. *Empirical vs. conceptual vs. theoretical observables.* One must guard against an oversimplified or too dogmatic a view of the concept of observable. Whether certain analytical expressions occurring in theoretical physics are in principle capable of empirical measurement can be quite a difficult and controversial question. Nevertheless, many of these expressions would qualify as conceptual observables—among them, e.g., the electromagnetic field strength averaged in the fashion described by Bohr and Rosenfeld, which ingenuity and technical advance may possibly render observable, although at this time no direct observation has been made or proposed. It is difficult to imagine a

useful quantum field theory which does not incorporate such conceptual observables in its formalism, in some way; the present chapters take a relatively conservative stand in assuming only that a certain physically restricted class of bounded functions of such field averages are conceptual observables, as we shall indicate more specifically later.

There are other complications, notably the indirect way in which the theoretical counterparts to certain important empirical observables (the S-matrix, the energy, and other generators of the fundamental symmetry group commuting with the energy) come to have interpretations as concrete operators, being on a somewhat different footing from the explicit functions of field averages. This does not alter the essential validity of the foregoing postulates, signifying only that they need to be supplemented by kinematical, dynamical, and statistical considerations. These will emerge later in our treatment of quantum fields; for brevity we must forego the consideration of such more general aspects as are accessible in a general physical system.

2. *Relative priority of states and observables.* Whether observables or states are more fundamental is somewhat parallel to the same question for chickens and eggs. Leaving aside metaphysics, either notion has certain distinctive advantages as a foundational concept, but no analytical treatment starting from the states exists as yet which is of the same order of comprehensiveness and applicability as that starting from the observables. In particular, the work of Birkhoff and von Neumann (1936) and of Mackey (1957), in which the states play the fundamental role, has not yet been developed to the point where their serviceability as possible frameworks for quantum field phenomenology is apparent.

3. *Operator algebras and states.* An important technical advantage of the observables as a starting point for phenomenology is that the theory of operator algebras, which has reached a point of considerable cogency after a quarter century of intensive development, can be brought directly to bear, at least with the present formulation of the concept.

The first clear-cut work in this direction is that of von Neumann, culminating in his 1936 paper, while on the purely mathematical side Stone's work on representation and commutative spectral theory in the late 30s and early 40s was particularly stimulating and relevant. A striking result which followed (1943) was Gelfand and Neumark's abstract algebraic characterization of C^*-algebra. While this result is expendable from a physical standpoint, the paper was influential in encouraging the study of C^*-algebras, which have turned out to be more useful for quantum phenomenology than the weakly closed rings whose study was initiated by von Neumann and Murray. While the distinction between an algebra of operators that is uniformly closed and one that is weakly closed may seem rather technical, it turns out to be vital, as is not so surprising when one considers that uniform convergence of observables has a direct physical interpretation, while weak convergence has only analytical significance.

An aspect of the theory of C^*-algebras which is important both mathematically and in relation to quantum mechanics is the duality between states and representations. We shall have occasion to describe and use later the mutual correspondence between states and representations developed in a relevant form by Segal (1947a).

References to Chapter I

The original articles of von Neumann (1927) are in some ways the best account of the origin of the material in this chapter, although (1932) subsumes most of these earlier papers. The paper of Segal (1947b) on which this chapter is largely based brought together some of the ideas of von Neumann (1936) with those of the representation theory of Stone, Gelfand, and others. The work of Lowdenslager and Sherman develops aspects of Segal's approach. An open question remaining here is that of the existence of "simple" infinite-dimensional Jordan algebras satisfying the given postulates which are not "special," i.e. essentially derivable from an associative algebra.

For associative systems the relevant theory is presented by Segal (1947a), which includes some material on the notion of maximal observation due to Dirac and its relation to Weyl's notion (1927) of pure state.

The extensive literature on C^*-algebras and on W^*-algebras (= "von Neumann algebras," or "rings of operators") is also relevant in a general way. Especially noteworthy, among the topics not already mentioned, is the theory of direct integrals of Hilbert spaces, which while in practice usually avoidable, is important as background material.

CHAPTER II

Canonical Quantization

Quantum mechanics of systems with a finite number of degrees of freedom is often expressed in terms of a finite set $p_1, q_1; p_2, q_2; \cdots; p_n, q_n$, of mutually commuting canonical pairs. A change in the frame of reference will change these into a new set of n such pairs. To deal in a compact and theoretically convenient fashion with all the fundamental "canonical variables," it is desirable to reformulate the notion of canonical system slightly. This reformulation turns out to be helpful in making the transition to a system of an infinite number of degrees of freedom as well as in clarifying the general notion of quantization.

We start from a finite-dimensional real linear vector space L whose physical interpretation is the classical physical (configuration) space associated with the system under consideration. (Thus for a system of n particles in 3-dimensional Euclidean space, $\mathscr{L}$ would be a $3n$-dimensional space.) The contragredient or dual space to $\mathscr{L}$, i.e. the space of all linear functionals on $\mathscr{L}$, will be denoted by $\mathscr{L}^*$. *A quantum-mechanical canonical system over* $\mathscr{L}$ may be defined in a purely mathematical manner, as a pair of unitary representations U and V of the additive groups of $\mathscr{L}$ and $\mathscr{L}^*$ respectively on a Hilbert space, which are weakly continuous and satisfy the Weyl relations:

$$U(x)V(f) = e^{if(x)}V(f)U(x) \qquad (x \in L, f \in \mathscr{L}^*).$$

For any such canonical system there exist by Stone's Theorem self-adjoint operators $P(x)$ and $Q(f)$ such that

$$U(tx) = \exp(itP(x)), \qquad V(tf) = \exp(itQ(f)).$$

These operators are called the canonical variables of the system. If $e_1, e_2, \cdots, e_n$ is a basis for $\mathscr{L}$ and $f_1, f_2, \cdots, f_n$ is a dual basis for $\mathscr{L}^*$ (i.e. $f_j(x_k) = \delta_{jk}$), then

$$\{P(e_1), Q(f_1)\}, \{P(e_2), Q(f_2)\}, \cdots, \{P(e_n), Q(f_n)\}$$

is a sequence of mutually commuting canonical pairs.

The existence of a canonical system over $\mathscr{L}$ is clear from the existence of the Schrödinger representation. Specifically, let $\mathscr{H}$ be the Hilbert space of all complex-valued functions that are square-integrable with respect to Lebesgue measure on $\mathscr{L}$ (which is unique within an inessential constant factor) and define

$$U(x): F(y) \to F(y + x),$$
$$V(f): F(y) \to e^{if(y)}F(y),$$

F being arbitrary in $\mathscr{H}$. It is straightforward to verify that the Weyl relations are satisfied.

Now it is a remarkable and very convenient circumstance (if conceivably physically expendable at the cost of considerable technical complication) that

ESSENTIAL UNIQUENESS OF CANONICAL SYSTEMS (STONE-VON NEUMANN): *Any canonical system over a finite-dimensional linear configuration space is, within unitary equivalence, a (discrete) direct sum of copies of the Schrödinger system.*

In other words, for an arbitrary canonical system (U', V') acting on a Hilbert space $\mathscr{H}'$, there is a decomposition

$$\mathscr{H}' = \mathscr{H}_1 \oplus \mathscr{H}_2 \oplus \cdots$$

of $\mathscr{H}'$ as a direct sum of subspaces $\mathscr{H}_1, \mathscr{H}_2, \cdots$ each of which is invariant under all the operators of the canonical system; and for each $\mathscr{H}_i$ there is a unitary transformation W_i of $\mathscr{H}_i$ onto the Schrödinger representation space $\mathscr{H}$ of square-integrable functions over $\mathscr{L}$, carrying the restriction of (U', V') to $\mathscr{H}_i$ onto the Schrödinger (U, V):

$$U(x)W_i = W_iU'(x),$$
$$V(f)W_i = W_iV'(f), \qquad (i = 1, 2, \cdots).$$

This is a technical mathematical result; its main physical import is that any two quantum canonical systems over the same configuration space $\mathscr{L}$, irreducible or not, are physically identical. For it follows from the Stone-von Neumann Theorem that the observables associated with the two systems are in algebraic isomorphism—this is true whether the full rings of operators generated by the canonical systems are employed or only the more manifestly physical C^*-algebra of smooth operators associated with the system. This is true in the case of the C^*-algebra mainly because the formation of a direct sum of copies of an operator does not affect its bound. In the case of the ring of operators there is algebraic uniqueness despite the fact that the weak topology may be affected by the formation of direct sums of copies, as may be shown by a special analysis.

This represents in a way a comprehensive mathematical explication of the physical equivalence of the Heisenberg and Schrödinger formulations of quantum mechanics. In these chapters it will play a role in simplifying the treatment of the quantization of Bose-Einstein fields. It makes it possible to speak simply of THE quantum-mechanical canonical system over a given configuration space $\mathscr{L}$.

Now in the foregoing treatment the canonical P's on the one hand and the canonical Q's on the other were treated quite separately, although it is clear that there is a certain symmetry between them. This distinction between the P's and the Q's has a clear physical basis in the case of systems of a finite number of degrees of freedom, as well as in non-relativistic systems more generally, but is devoid of relativistic significance in the case of quantum fields. In anticipation of the treatment of fields, it is useful to reformulate the treatment of finite systems in a way that involves no sharp distinction between the P's and Q's; and this reformulation leads to some incidental mathematical simplifications.

Let $\mathscr{M}$ denote the direct sum $\mathscr{L} \oplus \mathscr{L}^*$ of $\mathscr{L}$ with $\mathscr{L}^*$; i.e. $\mathscr{M}$ is a linear vector space a general element of which has two components (u, f) with u in $\mathscr{L}$ and f in $\mathscr{L}^*$, addition, etc. being componentwise. $\mathscr{M}$ is then the physical *phase space*. Let $P(\cdot)$, $Q(\cdot)$, denote a canonical system over L. For a general vector z in $\mathscr{M}$ of the form $z = (u, f)$, define a self-adjoint operator $R(z)$ by Stone's Theorem as the self-adjoint generator of the continuous one-parameter unitary group $\{S(t),\ -\infty < t < \infty$, where $S(t) = \exp[itP(u)] \exp[itQ(f)] \cdot \exp[-it^2 f(u)/2]\}$. Applying both sides of this equation to a vector in the domains of both $P(u)$ and $Q(f)$, differentiating with respect to t, and then setting $t = 0$, it results that $R(z) \subset P(u) + Q(f)$. (In fact $R(z)$ might have been defined as the closure of $P(u) + Q(f)$, but this is mathematically relatively inconvenient.) It follows without difficulty from the definition of $R(z)$ that

$$e^{iR(z)} e^{iR(z')} = e^{iB(z,z')/2} e^{iR(z+z')}, \quad (z' \text{ arbitrary})$$

where

$$B(z, z') = f'(u) - f(u').$$

Conversely, from this equation follows the Weyl relations, making suitable substitutions for z and z'. Thus a canonical system over a configuration space $\mathscr{L}$ may be alternatively defined as a map $z \to R(z)$ from the associated phase space $\mathscr{M}$, satisfying the given relation, i.e. $R(z)$ is for each z a self-adjoint operator in a Hilbert space, the generalized Weyl relation given above is satisfied, and in addition $e^{iR(z)}$ is a weakly continuous operator function of the variable z.

In infinitesimal form, this form of the Weyl relation is

$$[R(z), R(z')] \subset -iB(z, z').$$

It is helpful to note also that $R(z)$ depends linearly on z, insofar as its unbounded character permits:

$$R(z + z') \subset R(z) + R(z'),$$
$$R(az) = aR(z) \quad (a \text{ an arbitrary non-vanishing real number}).$$

It is sometimes convenient, especially in field theory, to introduce complex canonical variables, which depend actually not only in a real-linear but also complex-linear fashion on the vector z, which designates the position in phase space. To arrive at a formulation of phase space as a complex rather than real-linear manifold, we consider that frequently the configuration space $\mathscr{L}$ is given not merely with the structure of a real-linear vector space, but also is endowed with a distinguished metric, e.g. is given as euclidean space, frequently. We suppose then that there is given on $\mathscr{L}$ a distinguished real positive definite symmetric bilinear form (x, y) in the vectors x and y of $\mathscr{L}$.

In this event the dual $\mathscr{L}^*$ of $\mathscr{L}$ may be canonically mapped onto $\mathscr{L}$: if f is in $\mathscr{L}^*$, there will be a unique element u on $\mathscr{L}$ such that

$$f(x) = (x, u) \quad (x \text{ arbitrary in } \mathscr{L}).$$

Writing $f = u^*$ and $f^* = u$, then it is clear that $u^{**} = u$, etc. A general element z of $\mathscr{M} = \mathscr{L} \oplus \mathscr{L}^*$ then has the form

$$z = u \oplus v^*$$

for some u and v in $\mathscr{L}$. Now define an operation j on $\mathscr{M}$ as follows:

$$jz = -v \oplus u^*.$$

It is readily seen that j is a real-linear transformation $[j(az) = aj(z)$ for a real, and $j(z + z') = jz + jz'$ for arbitrary z and z' in $\mathscr{M}]$, and that $j^2 = -1$, 1 denoting here the identity transformation. Thus j behaves much like multiplication by the complex number i, and it is actually a well-established and rather familiar fact that one may now, as a matter of pure algebra, introduce a complex structure in $\mathscr{M}$ by defining the action of a general complex number $a + ib$ as follows:

$$(a + ib)z = az + bjz.$$

There is no difficulty in verifying that with this definition, $\mathscr{M}$ becomes a linear vector space over the complex number field.

Now define for any z in $\mathscr{M}$ an operator $C_0(z)$ by the equation

$$C_0(z) = [R(z) - iR(iz)]/2^{1/2}$$

and put $C(z)$ for the closure of $C_0(z)$, as can be shown to exist. It can then be shown that $C(z)$ is essentially a complex-linear function of z, in contrast to the merely real-linear character of $R(z)$, notably, $C(az) = aC(z)$, a an arbitrary non-vanishing complex number. In addition to this simple algebraic property, the $C(z)$ have somewhat more convenient commutation relations than the $R(z)$, from an infinitesimal point of view. To formulate these, it is useful to note that $\mathscr{M}$ not only has a distinguished complex structure but actually a distinguished unitary space structure. Defining $S(z, z') = -B(iz, z')$, it is easily verified that $S(z, z') = (u', u) + (v, v')$, so that $S(z, z')$ is a positive definite real symmetric form on $\mathscr{M}$. It is now straightforward to verify that the definition

$$\langle z, z' \rangle = S(z, z') + iB(z, z')$$

yields a positive definite inner product on $\mathscr{M}$, which together with the complex structure on it, renders $\mathscr{M}$ a unitary space. The commutation relations of the $C(z)$ and their Hilbert space adjoints may now be stated:

$$[C(z), C(z')] = 0, \qquad [C(z), C(z')^*] \subseteq -\langle z, z' \rangle.$$

As the simplicity of the relations involving $C(z)$ might suggest, the so-called creation and annihilation operators $C(z)$ and $C(z)^*$ (the reasons for these designations will appear later in connection with the particle interpretation for field theory) are very useful in algebraic analysis of canonical systems, of polynomial expressions in the canonical variables, etc. On the other hand, these operators are unbounded and very far from being normal (or even diagonalizable), and so are troublesome to deal with in a rigorous analytical way. For this reason, and since we are here more concerned with foundations than with computations, we shall generally avoid the use of the $C(z)$.

The main significance of the unitary phase space $\mathscr{M}$ will only appear later, in connection with infinite-dimensional systems. It will turn out to insure the existence of a vacuum state for the associated quantum system.

It ought to be noted that there is no physical necessity for a classical phase space $\mathcal{M}$ to be a *linear* manifold; this is merely the simplest and mathematically fundamental case. Intuitively the configuration space $\mathcal{L}$ could conceivably be virtually an arbitrary smooth manifold. How "quantum conditions" may appropriately be introduced for a non-linear phase space is a natural question at this point, from a logical viewpoint, but it is convenient to postpone its discussion until the quantization problem for non-linear fields is taken up later.

Now so far the configuration space $\mathcal{L}$ and phase space $\mathcal{M}$ have been finite-dimensional. What difficulties, if any, appear in the case of an infinite-dimensional system? The field-theoretic systems which are our main concern involve in fact infinitely many degrees of freedom.

There is no difficulty in making a formal extension of the basic concepts involved in canonical systems. If $\mathcal{M}$ is an infinite-dimensional linear space (the phase space of a classical field, say) with a distinguished skew-symmetric bilinear form B, the notion of canonical system may be defined just as above. A mathematician will immediately raise the question of the existence and uniqueness of such; and these turn out to involve non-trivial mathematical developments.

Physically new difficulties appear in an infinite-dimensional system in connection with dynamics. The rather celebrated divergences of field theory represent troubles with the formulation of dynamics, which have persisted in theoretical physics ever since the original work of Dirac on quantum fields. Fairly recent mathematical analysis has indicated that these divergences are closely connected with the mathematical developments relevant to the existence and uniqueness of canonical systems.

Before going into the infinite-dimensional case in detail, it may be well to develop rapidly a fairly concrete example of what may be involved. We shall present one which is about as representative as its relatively elementary and special form permits. First, however, it is necessary to dispose of the general phenomenological question of the formulation of dynamics.

In the older conventional theoretical physical treatment of quantum mechanics, it was essentially taken for granted that a dynamical transformation was effected by a unitary (or in a few quite special cases, anti-unitary) transformation. Such a transformation U transforms an observable (represented by the operator) X into the observable $U^{-1}XU$, or the wave function ψ into the wave function $U\psi$, and the situation seemed fundamentally quite simple. More generally, the action of time on the system (i.e. its temporal development) was given by a one-parameter family U_t of such transformations, so that the observable $X(t)$ at time t in question was given by the equation

$$X(t) = U_t^{-1}X(0)U_t.$$

In the case of a system acted upon by no external forces except time-independent ones, the family U_t was a one-parameter group: $U_tU_{t'} = U_{t+t'}$.

From the more conservative point of view of the phenomenology developed in the first chapter, this is a rather ad hoc formulation of dynamics. On a purely physical, rather than academic analytical, basis, the main point about a dynamical transformation is that it preserves the physical algebra of observables. More specifically, if

$$X \to X', \quad \text{and} \quad Y \to Y',$$

then

$$X + Y \to X' + Y', \quad X^2 \to X'^2, \quad \text{and} \quad aX \to aX' \qquad (a \text{ real}).$$

In the particular case that these observables are represented by operators and X' has the form $X' = UXU^{-1}$, U being unitary, it is evident that these conditions are satisfied, but quite conceivably, at least for someone with a background in pure mathematics, there were other cases as well.

A physicist most of whose work has been along computational lines is apt to presume that the foregoing definition of a dynamical transformation—as an *automorphism* of the algebra of observables, to use the precise mathematical term—is analytically equivalent to the definition in terms of unitary or anti-unitary operators. This happens to be actually the case for the algebra of all bounded self-adjoint operators on a Hilbert space, as is not difficult to establish rigorously. Every automorphism of this system is implementable by some unitary or anti-unitary operator—in fact a unique one, it is easily seen, within multiplication by a scalar. This reduces the question to that of whether the system of all bounded self-adjoint operators on a Hilbert space is actually the right one.

Now there is no apparent physical reason why this should be the case, even in elementary quantum mechanics. It is usually assumed for mathematical convenience and because it does no visible harm. But it must be emphasized that it has no empirical basis, and so should be discarded as soon as it is found to lead to analytical difficulties. It is fairly evident at this point that it does no great harm in the case of a canonical system with a finite number of degrees of freedom to assume that every bounded self-adjoint operator is observable. The Stone-von Neumann uniqueness theorem implies that, no matter how the canonical system is represented by operators in a Hilbert space, every self-adjoint bounded operator will have a well-defined expectation value in every regular state, and so will generally behave much like an observable. The only visible trouble is that not all states, in the abstract sense of von Neumann's characterization, appear to be truly physical. However, if one is maximally strict and considers only the smooth observables described in the first lecture, the algebra of which has precisely those states which are considered to be the empirically accessible ones, the situation is not materially affected. The algebra of smooth observables is in fact in the Schrödinger representation the algebra of completely continuous operators on the Hilbert space, augmented by the identity operator; and every automorphism of the algebra of self-adjoint elements of this algebra is implementable by an (essentially unique, as before) unitary or anti-unitary transformation.

Thus it turns out finally that there is no material harm, in the case of a system of a finite number of degrees of freedom, in making the assumption that the observables are self-adjoint operators, the dynamical development is given by a one-parameter family of unitary operators, etc., because of a favorable technical situation. Thereby "physical intuition" is vindicated, as regards finite systems. But in the case of infinite systems, which went mathematically unexamined for many years, the corresponding physical folk theorems have turned out to be completely false, as established by fully precise mathematical work. Notably it is not true, as had always been implicitly assumed, that every or even most dynamical transformations can be implemented by unitary or anti-unitary operators (when the observables are represented by operators in the apparently most relevant representation). This is not true for any complicated technical reasons analogous to the non-differentiability of certain functions, such as might be excludable on physical grounds like the non-occurrence of similar pathological situations in nature, but developed rather out of quite simple analytical reasons, fundamentally. In fact, it seems probable that in field theory it is only the kinematical transformations, and virtually never the crucial ones giving the physical temporal development of the system, that can be implemented by unitary transformations, when the observables are represented in any explicit fashion (e.g. as operators on a "bare," or alternatively "free physical" field—cf. below).

To examine a particular case, let us first set up, quickly although non-covariantly, a sequence of canonical variables $p_1, q_1, p_2, q_2, \cdots$.

Let R denote the system of all real numbers, and let m be the measure whose element is $(\pi)^{-(1/2)}\exp(-x^2)\,dx$. Let S denote the measure space consisting of the infinite direct product of a countable set of copies of the probability space (R, m), and let $\mathscr{H}$ be the Hilbert space of all complex-valued square-integrable functions over this space. A general point in S has the form $(x_1, x_2, \cdots)$, where the x_k are real numbers, and a general element f of $\mathscr{H}$ will be a function $f(x_1, x_2, \cdots)$ of this countable set of real variables. Now let $U_k(s)$ denote, for any real s, the following unitary transformation on $\mathscr{H}$:

$$U_k(s): f(x_1, \cdots, x_k, \cdots) \to f(x_1, \cdots, x_k + s, \cdots)\exp[-(2xs + s^2)/2].$$

For any real t, let $V_k(t)$ denote the following unitary transformation on $\mathscr{H}$:

$$V_k(t): f(x_1, \cdots, x_k, \cdots) \to f(x_1, \cdots, x_k, \cdots)e^{itx_k}.$$

It is straightforward to verify that the Weyl relations are satisfied:

$$\begin{aligned} U_k(s + t) &= U_k(s)U_k(t), \\ V_k(s)V_k(t) &= V_k(s + t), \\ U_k(s)V_k(t) &= e^{ist\delta_{kk'}}V_{k'}(t)U_k(s). \end{aligned}$$

Thus we have constructed an infinite canonical system, the self-adjoint canonical variables being conveniently definable as the self-adjoint generators of the one parameter groups, say p_k generates $[U_k(t);\ -\infty < t < \infty]$, and q_k generates $[V_k(t);\ -\infty < t < \infty]$. (Incidentally, these p_k and q_k differ by factors of $2^{\pm(1/2)}$ from a certain standard set of canonical variables, which may be called the zero-interaction ones, which will be described later.)

Now it is frequently convenient in quantum theory to go over from one set of canonical variables to another (on the same Hilbert space) satisfying the same commutation relations. According to an old physical folk theorem any two such canonical systems were related by a unitary transformation, so there was no harm in making such a change of variables. Actually, no explicit Hilbert space on which the canonical variables acted as self-adjoint operators was ever mentioned. This made the precise meaning of the theorem elusive, and also made it impossible to give a counter example to it without first giving a more precise formulation. However, an explicit Hilbert space has been set up above, so that it is a perfectly definite mathematical question whether two given canonical systems, say

$$p_1, q_1, p_2, q_2, \cdots \quad \text{and} \quad p_1', q_1', p_2', q_2', \cdots,$$

are connected by a unitary transformation; i.e. whether there exists a unitary transformation U on the Hilbert space such that

$$Up_kU^{-1} = p'_k, \qquad Uq_kU^{-1} = q'_k \qquad (k = 1, 2, \cdots).$$

The most popular transformations are those which are linear inhomogeneous in the p's and q's. Let us consider the maximally simple case of a "scale" transformation:

$$p'_k = cp_k, \qquad q'_k = c^{-1}q_k \qquad (c > 0).$$

Now it is mathematically demonstrable that it is only in the trivial case $c = 1$ that this transformation is unitarily implementable. As a matter of fact, no transformation of the form $p_k \to f(p_k), q_k \to g(q_k)$ $(k = 1, 2, \cdots)$ is unitarily implementable except the trivial ones with $f(x) = \pm x$ and $g(x) = \pm x$. And if the transformation varies with k, say

$$p'_k = c_kp_k, \qquad q'_k = c_k^{-1}q_k \qquad (c_k > 0),$$

then it is necessary and sufficient for unitary implementability that the following infinite product be convergent:

$$\prod_k (2c_k/(1 + c_k^2)),$$

a condition which requires in particular that $c_k \to 1$, and thereby excludes all of the actual cases in which the transformation has been used in practice (cf. e.g. von Neumann's book for an example of this use).

To gain some insight into the nature of the field-theoretic divergences, let us examine the scale transformation given above more fully, in relation to the older conventional practice. The existence of an implementing unitary transformation seemed clear particularly because it could be written down explicitly:

$$U = \exp\left[ig \sum_{k=1}^{\infty} (P_kQ_k + Q_kP_k)\right],$$

where g is a constant depending on c. To check formally that U has, at least in a figurative sense, the stated property reduces to a problem in one dimension that it is actually not difficult to resolve in an analytically precise way. But in an infinite number of dimensions there simply is no such operator: the sequence of unitary operators, $[\exp(ig\sum_1^n(P_kQ_k + Q_kP_k)); n = 1, 2, \cdots]$ converges, but not to a unitary operator; the limit, in fact, is 0.

The figurative operator U provides an example of an essentially "divergent" operator. The really interesting operators of field theory are, first of all, those describing the temporal development of the system. In the so-called "interaction" representation, these are given in a fairly explicit form which in the case of the transition from time t to time $t + dt$ are fairly similar to U, the main difference being that the expression in the canonical operators in the exponential is usually cubic or linear rather than quadratic. There is every indication that these formally unitary operators giving the temporal development fail to exist in any simple and effective mathematical sense, although in view of their heuristic formulation it is impossible in the nature of things to prove mathematically that this is the case. Nevertheless these "divergent" operators may be used in computations to give accurate numerical results, provided suitable "renormalization" is effected. How such renormalization may remove apparent infinities may be seen in the case of the present operator U. Even the matrix elements (Uf, g) of U between the simplest state vectors f and g work out as divergent—they vanish identically, although these matrix elements should in some sense combine to give a unitary matrix. However, if the (Uf, g) are multiplied by a suitable "infinite constant," a perfectly finite, numerically computable, matrix is obtained. Specifically, let $a = (e^{i\alpha(PQ+QP)}1, 1)$, where 1 denotes the function identically unity on the real line, the inner product and P and Q are formulated as above except that only one copy, rather than a countable set of copies, of the real line with the designated probability measure is used. There is no difficulty in computing a explicitly, but its precise value is not relevant here. It is not at all difficult to see, and this is the point, that although $(U_k f, g) \to 0$, where $U_n = \exp[ig\sum_1^n(P_kQ_k + Q_kP_k)]$,

$$\lim_{k\to\infty} a^{-k}(U_k f, g)$$

exists for a dense set of f and g, say for all those of the form $\prod_k f_k(x_k)$, where all of the f_k except a finite number are identically unity, the others being polynomials.

Thus, in a figurative sense, if the "infinite constant" $a^{-\infty}$ is designated Z, then the renormalized matrix elements $Z(Uf, g)$ are finite and well-defined. If we are concerned only with the comparison of the matrix elements of U with those of another divergent operator which can however be renormalized by multiplication by the "same" Z, then it is clear that we have no difficulty. It is also clear that this depends on having an especially felicitous situation, and is very far from beginning to resolve the foundational questions involved.

Notes to Chapter II

1. *Fermi-Dirac quantization.* Why the relations $[p_j, q_k] = i\delta_{jk}$? Originally, Dirac introduced these rules simply by analogy with those introduced by Heisenberg. In a way it is extraordinary that they were so successful for although formally similar they have quite a different physical role from the Heisenberg rules. The latter, to a mathematician concerned with physics and at a suitable level of sophistication, are a way of describing the geometry of a single non-relativistic electron. They assert roughly that its ambient space is three-dimensional euclidean space, and that the action of the group of translations on this space is rather fundamental, in a fashion in which Planck's constant intervenes very materially. The Dirac field quantization rules, on the other hand, state nothing about the geometry of any single particle, being entirely independent of special assumptions as to the description of the particle, except essentially that its states can be represented by vectors in an infinite-dimensional linear space. They describe rather a possible mode of independent but structured existence of an indefinite number of identical particles, without reference to the species of the particle, in particular independently of whether the particle has a three-dimensional ambient space, an infinite-dimensional one, or none at all (the linear vector space representing the states being possibly of quite a different character from a collection of functions on some type of manifold). Moreover, the cited rules do not apply to the familiar material particle, the electron; rather, they were introduced to deal with photons.

The historical development of the rules for dealing with electrons is a rather lengthy story, but an algebraist might be led to them by contemplating the rules

$$[R(z), R(z')] = -iB(z, z'),$$

where $B(z, z')$ is a scalar. Necessarily then $B(z, z')$ is skew-symmetric in z and z', and if the $R(z)$ are to be hermitian, B must have real values; and for a fully non-trivial theory, B must be non-degenerate. Conversely, if B has these properties, a coherent and effective mathematical theory is possible, as indicated more fully below.

It should be natural for an algebraist to inquire about the replacement of the commutator $[R(z), R(z')]$ by the anti-commutator, $[\cdot, \cdot]_+$, where $[A, B]_+$ is defined as $AB + BA$, yielding the rules

$$[R(z), R(z')]_+ = S(z, z'),$$

where the values of S are scalars. Necessarily then $S(z, z')$ is symmetric in z and z', and if the $R(z)$ are to be hermitian, S must have real values. Again, for full non-triviality, S must be non-degenerate. It then develops that a theory quite different from but as structured as that involving the commutators can be developed and that, for essentially familiar quasi-empirical reasons, the resulting algebra is suitable for describing an indefinite number of electrons (when the underlying linear space and symmetric form are suitably chosen).

There are some really substantial differences between the anti-commutation and the commutation rule algebras (the so-called Fermi-Dirac and Bose-Einstein quantization rules, respectively). The anti-commutation rules are simpler in the analytical respects with which we are primarily concerned in that the operators $R(z)$ are bounded, but are more complicated in some algebraic respects, notably in the lack of combinations of the field variables $R(z)$ generating a maximal abelian system, in the most relevant representation. However most of the truly fundamental features of quantum field theory seem to be representable either in terms of Bose-Einstein or Fermi-Dirac fields. For simplicity and brevity, as well as for their general relevance, these chapters will deal primarily with Bose-Einstein fields, whenever only one type of field is being considered. The typical situation where both types are involved, the trilinear (linear-bilinear) interaction between a Bose-Einstein and a Fermi-Dirac field, which is the most important single type and includes for example quantum electrodynamics, will be discussed later.

There are two possible answers to the natural question as to other possible modes of quantization, the empirical and the theoretical. Empirically all elementary particles are perfectly well represented, as regards their statistics and as far as they can be observed, by one of the two types described. In view of the limited number of such particles known and the difficulty of closely examining some of them, this is not too conclusive. But also theoretically there are results indicating that no other systems of equal analytical simplicity and coherence exist. However, there seems to be no clear-cut consensus of opinion on the extent to which such results are conclusive.

It might indeed conceivably be of mathematical interest to explore the possible existence of free fields satisfying trilinear or higher order pseudo-commutation relations and other such desiderata of field theory. The indications are that no such schemes exist which are applicable to an arbitrary Hilbert space of single-particle states, but this does not at all rule out the possible existence of such schemes which are Lorentz-invariant and specially adapted to spaces of (spin-space-valued) functions on conventional space-time.

2. *Singular unitary transformations on fields and singular transformations in Wiener space.* The non-existence of a unitary transformation which induces the pseudo-canonical transformation

$$q_k \to cq_k, \qquad p_k \to c^{-1}p_k \quad (c^2 \neq 1) \quad (k = 1, 2, \cdots)$$

where the canonical variables are in the representation described above, which is one of the simplest violations of the familiar folk theorem concerning such transformations, is related to a singular analytical circumstance which was noted several years before the quantum-theoretic situation was explored. This is the result of Cameron and Martin (1947) that the simple transformations $x(t) \to cx(t)$ ($c > 0$, $c \neq 1$) on Wiener space have highly pathological measure-theoretic properties. It may be said heuristically that *if* there did exist a unitary transformation implementing the pseudo-canonical transformation described, or even only the maps $q_k \to cq_k$ ($k = 1, 2, \cdots$), then the transformation described in Wiener space would be absolutely continuous.

To sketch the connection briefly, the Brownian motion process $x(t)$, say for $0 \leqq t \leqq 1$, may be represented as a Fourier series

$$x'(t) \sim \sum_k q_k e^{2\pi ikt},$$

where the q_k are independently distributed normal random variables of unit variance. It is evident that the transformations $x(t) \to cx(t)$ and $q_k \to cq_k$ correspond. Now if the former transformation were absolutely continuous, the transformation

$$F[x(\cdot)] \to F[cx(\cdot)]\,(dw_c/dw)^{(1/2)}$$

would be unitary (here w represents Wiener measure and w_c its transform under the indicated operation). It is not difficult to check that this unitary operator would transform the operation of multiplication by q_k into that of multiplication by cq_k ($k = 1, 2, \cdots$). Thus, if the scale transformation were absolutely continuous, there would exist a unitary transformation transforming the q_k into the cq_k. On the other hand, if the transformation $q_k \to cq_k$ ($k = 1, 2, \cdots$) were implementable by a unitary operator, it can be shown that the corresponding transformation in the infinite product space described above would be absolutely continuous and so also would be the scale transformation in Wiener space.

In general the absolute continuity of a transformation in Wiener space is readily seen to imply the existence of a unitary operator implementing a corresponding transformation of the canonical variables. The converse is true as regards transformation of the q_k alone, although less obvious. Naturally the possibility of transforming the q's in a stated manner implies in general nothing regarding the possibility of transforming the p's in a stated manner. The canonical p's do not arise in connection with the theory of Brownian motion, and while an extremely simple condition can be given in many cases for the unitary implementability of a given transformation of the p's and q's, it does not have an especially simple interpretation in terms of Wiener space.

References to Chapter II

The only published mathematical proofs, at the present time, of the non-unitary-implementability of certain apparently canonical transformations of Weyl systems are provided by Segal's general criteria (1958). What amounts to a proof of the non-unitary-implementability of some special pseudo-canonical transformations in the Fermi-Dirac case is implicit in the material near the end of a paper of von Neumann (1938), which, while directed towards applications to quantum fields, does not explicitly deal with field variables or their representation. Particular examples in the same general direction, of a mathematically partially heuristic character, were given by Friedrichs (1953), Haag (1955), and Schweber and Wightman (1955). Rigorous results on the classification of Weyl systems were announced by Gårding and Wightman (1954), but the proofs have not yet been published. These authors employ an "occupation number" representation which has a somewhat obscure physical interpretation and is dependent on the choice of a basis in the single-particle space.

Our first observation of the existence of inequivalent representations of the commutation relations (1950) appears to provide one of the simplest rigorous demonstrations of the phenomenon in relevant concrete terms. Any unitary transformation taking $q_k \to c_k q_k (k = 1, 2, \cdots)$ also takes $\lim_n \sum_{k<n} a_k q_k$ into $\lim_n \sum_{k<n} a_k c_k q_k$. From the interpretation of the canonical variables in the standard representation in terms of a sequence of identically distributed normal random variables, and the familiar simple results concerning the sums of independently distributed random variables, such a limit as $\lim_n \sum_{k<n} b_k q_k$, the b_k being real numbers, will exist on a dense domain if and only if the series $\sum_k b_k^2$ is convergent. Obviously, this condition is not invariant under

the transformation $b_k \to c_k b_k$, in general. An elaboration of this argument showed that a number of realistic interaction hamiltonians in quantum field theory could be transformed into densely defined operators in a Hilbert space by an adaptation of the indicated transformation, but in continuum inequivalent ways. This apparent lack of uniqueness gave the result a seemingly negative character, until the development of the representation-independent formalism described in Chapters IV and V indicated that the derived physical dynamics might nevertheless well be essentially unique (cf. Chapter VII).

The foregoing results were communicated orally to van Hove, among others, who not long afterwards formulated his well-known paradox (1952), to the effect that the eigenstates of the "free" hamiltonian

$$H_0 = \sum_k (a_k p_k^2 + b_k q_k^2)$$

are orthogonal to those of the "total" hamiltonian

$$H = H_0 + H_I,\ H_I = \sum_k (c_k p_k + d q_k),$$

for suitable constants a_k, b_k, c_k, d_k $(k = 1, 2, \cdots)$. This of course contradicts the standard notion that the eigenstates of either form a complete orthonormal set. This is necessarily quite a heuristic situation: what H_0 and H_I mean as operators, in what Hilbert space they operate, cannot both really be specified. The paradox has since come to be thought of as being related to the unitary inequivalence of the canonical variables p_k, q_k, and their linear transforms, obtained by completion of the square, in terms of which H has the same expression essentially as H_0. While these may be formulated as rigorous representations which are indeed inequivalent, relative to the relevant (free-field) representation for the original p's and q's, the Hilbert spaces on which both sets of canonical variables act are identical, so that the orthogonality of the eigenstates would seem to remain somewhat paradoxical. Actually, in terms of the representation-free formalism, the situation is quite transparent (cf. Chapter V).

CHAPTER III

Quantization and Relativistic Wave Equations

By "quantization" we mean the passage from a classical mechanical system to a corresponding quantum-mechanical one. For example, if the classical phase space is a linear space $\mathcal{M}$ of finite dimension, with fundamental non-degenerate skew-symmetric form $B(z, z')$ $(z, z' \in M)$, the corresponding quantized system is that whose algebra of observables is generated by the canonical variables $R(z)$, where these are operators in a Hilbert space satisfying the commutation relations

$$[R(z), R(z')] \subseteq -iB(z, z').$$

If $\mathcal{M}$ is formulated as the set of all $2n$-vectors $(p_1, p_2, \cdots, p_n, q_1, q_2, \cdots, q_n)$, and $B(z, z')$ in the form $\sum_k (p'_k q_k - p_k q'_k)$, as is always possible by a suitable choice of coordinates, and agrees with the conventional notation in classical mechanics, then the $P_k = R(e_k)$ and $Q_k = R(f_k)$, where the e_k and f_k make up the basis for $\mathcal{M}$ such that the designated vector equals $\sum_k (p_k e_k + q_k f_k)$, satisfy the conventional Heisenberg commutation relations. A similar definition would apply to the case when $\mathcal{M}$ is infinite-dimensional, but the existence and uniqueness of canonical variables of the type described is then not clear a priori. We intend to take up these questions, but before doing so it is probably well to indicate the physical motivation for the consideration of this case, and to explain the origin of the form $B(z, z')$. We begin therefore by considering one of the simplest non-trivial examples of a relevant physical situation, the so-called neutral scalar, or Klein-Gordon, field.

This field has a particular state of it determined by a solution ϕ of the equation

$$\Box\phi = m^2\phi,$$

where ϕ is a real-valued function of the real variables x_0, x_1, x_2, x_3, and $\Box$ denotes the differential operator $-\partial^2/\partial x_0^2 + \partial^2/\partial x_1^2 + \partial^2/\partial x_2^2 + \partial^2/\partial x_3^2$. Actually this is only a heuristic statement, for on the one hand we require a certain boundedness condition on ϕ to admit it as a physically realizable state, and in addition admit generalized functions as possible derivatives of the ϕ's. Although there is considerable point in identifying precisely this class of ϕ's by means of an analysis wholly within the physical space (with general point (x_0, x_1, x_2, x_3)), this analysis is much longer and more complicated than is required if we are willing to take advantage of the constant-coefficient character of the partial differential equation under consideration through the use of Fourier transforms. Since our main objective is to present a simple example of an interesting infinite-dimensional phase space rather than to indicate how to deal with variable-coefficient linear partial differential equations from the point of view of quantization, it is reasonable to use this expedient.

We therefore replace the consideration of ϕ by the consideration of what is, roughly, its Fourier transform, the function $\tilde{\phi}$ defined on the hyperboloid $k \cdot k = m^2$, where k denotes a vector (k_0, k_1, k_2, k_3) and $k \cdot k' = k_0k_0' - k_1k_1' - k_2k_2' - k_3k_3'$, such that

$$\phi(x) = \int_{k^2=m^2} e^{ix\cdot k}\tilde{\phi}(k)\frac{d_3k}{|k_0|},$$

where d_3k denotes $dk_1dk_2dk_3$; the indicated element of measure, $|k_0|^{-1}d_3k$, is simply that induced by conventional Lebesgue measure in the four-dimensional k-space. The reality of ϕ requires only that $\tilde{\phi}$ be complex-valued and satisfy in addition the equation:

$$\tilde{\phi}(-k) = \tilde{\phi}(k)^*,$$

where the * denotes the complex conjugate. The precise class $\mathscr{M}$ of $\tilde{\phi}$'s which the theory admits as representative of actual physical fields is then that of all such measurable $\tilde{\phi}$'s which are square-integrable on the hyperboloid with respect to the indicated measure. $\mathscr{M}$ is then evidently a real Hilbert space relative to the inner product

$$(\tilde{\phi}, \tilde{\psi}) = \int_{k^2=m^2} \tilde{\phi}(k)\tilde{\psi}(k)^* \frac{d_3k}{|k_0|},$$

which is always real.

To arrive at the fundamental skew-symmetric bilinear form on $\mathscr{M}$, it may be noted, rather remarkably, that $\mathscr{M}$ is in a natural way a *complex* Hilbert space. If j denotes the transformation $\tilde{\phi}(k) \to i\varepsilon(k)\tilde{\phi}(k)$, where $\varepsilon(k) = \pm 1$ according as k_0 is positive or negative, then it is easily verified that j is an orthogonal transformation on $\mathscr{M}$ such that $j^2 = -I$, where I denotes the identity transformation. From this it can be deduced in a purely algebraic way that $\mathscr{M}$ becomes a complex Hilbert space relative to the inner product $\langle\ ,\ \rangle$, where

$$\langle\phi, \psi\rangle = (\phi, \psi) - i(j\phi, \psi),$$

together with the definition of multiplication by complex numbers:

$$(a + ib)\phi = a\phi + bj\phi.$$

[Of course, $i\phi(k)$ exists as a function, but it is not a member of the class $\mathscr{M}$, and should not be confused with $j\phi(k)$. The action of j is on the members of $\mathscr{M}$ as entities, and not upon the functional values of representatives for the elements of $\mathscr{M}$.]

This complex structure, incidentally, is closely connected with the separation of the field into positive and negative frequency parts, as is commonly done in the theoretical physical literature. If we now set

$$B(\phi, \psi) = -(j\phi, \psi),$$

it is evident that B is a skew-symmetric form on $\mathscr{M}$.

The classical phase space $\mathscr{M}$, with fundamental skew form B, is then ripe for quantization. This depends on the existence and uniqueness of canonical systems in the infinite-dimensional case for a given classical linear system $(\mathscr{M}, B)$, a matter requiring substantial consideration, because of important differences from the finite-dimensional case. We leave this quite general question until the next chapter, and continue the examination of the specific phase space $\mathscr{M}$ determined by the Klein-Gordon equation.

In principle the equations of a field do not need to admit any symmetries but it is only when suitable symmetries exist that it is possible to have conservation laws for energy, momentum, etc., of the conventional type. The definition of energy, momentum, etc., in the general case depends upon the existence of a suitable treatment for the symmetrical case. For this and other reasons the invariance properties of the present system under the action of the Lorentz group are important.

A Lorentz transformation $T: x \to \Lambda x + a$, induces in an obvious manner a transformation of the functions ϕ defined on space-time:

$$\phi(x) \to \phi(\Lambda x + a).$$

It is straightforward to verify that this transformation, say T', commutes with the D'Alembertian $\Box$. It thus appears that the space $\mathscr{M}$ is left invariant by the action of the Lorentz group, but to be quite precise about it, the action of the group on functions of momentum should be given a definite formulation. The action

$$\tilde{\phi}(k) \rightarrow e^{i\Lambda^{-1}a\cdot k}\tilde{\phi}(\Lambda k)$$

is formally indicated and defines, as is easily seen, an orthogonal transformation on $\mathscr{M}$ as a real Hilbert space.

Now up to this point we have not placed any restriction on m^2, which could have been negative, so that m is pure imaginary. If we now restrict consideration to the case when m is real, it can be deduced that the foregoing orthogonal transformation commutes or anti-commutes with the transformation j defined above, according as the Lorentz transformation is time-preserving or time-reversing (see below). In particular, the proper Lorentz transformations are represented by unitary transformations on $\mathscr{M}$ as a complex Hilbert space. When m is not real, the transformations are definitely not unitary, a situation closely related to what is referred to in the theoretical physical literature as the impossibility of making a covariant separation of a field with imaginary mass into positive and negative frequency parts.

To examine more closely the character of the foregoing transformation for the case of the reversal operations of the Lorentz group, it is useful to recall that there exist two characters χ_s and χ_t of the full inhomogeneous Lorentz group G, mapping the group onto ± 1, and distinguished by the properties of taking space reversal, but not time reversal, resp. time reversal but not space reversal, into -1. The kernel of χ_t is a subgroup G_0 of G of index 2, the so-called orthochronous group. It is readily checked directly that the elements of G_0 are represented by unitary transformations. To see that the remaining elements of G are represented by "anti-unitary" transformations, i.e. one-to-one linear transformations of the complex Hilbert space onto itself such that $\langle\tilde{\phi}, \tilde{\psi}\rangle \rightarrow \langle\tilde{\psi}, \tilde{\phi}\rangle$, it suffices to note that space-time-reversal is represented simply by complex conjugation on $\mathscr{M}$, when represented as above by functions on momentum space. This comes about in a perfectly natural way, but people have sometimes tried to find some unitary transformation that could be made to correspond to time reversal; it can be shown that there is none which will preserve the important representation property of the transformations which we have defined.

That is to say, it is rather clear that if $U(T)$ denotes the transformation on $\mathscr{M}$ that is induced by the Lorentz transformation T^{-1} (the exponent -1 is required for mathematical convenience), then

$$U(T)U(T') = U(TT'),$$

for arbitrary Lorentz transformations T and T'. The map $T \to U(T)$ may be called a semi-unitary representation of G on $\mathcal{M}$; it is unitary actually on the orthochronous Lorentz group. It is in fact the best known irreducible unitary representation of this group. That it is a continuous representation, i.e. that $(U(T)\phi, \psi)$ is a continuous function of T, for any fixed ϕ and ψ, may be readily verified by standard methods. It is therefore one of the representations treated in Professor Mackey's chapters, to which we refer for the (non-trivial) proof of irreducibility. Once this is established, it is not difficult to go slightly further and show that the representation is real-irreducible; i.e. not only are there no non-trivial closed invariant subspaces of $\mathcal{M}$ as a complex Hilbert space, under the action of the orthochronous group, but the same is true for $\mathcal{M}$ as a real Hilbert space. From this it may be deduced, making use of an infinite-dimensional analogue to Schur's lemma, that the form B given above is the *only* continuous Lorentz-invariant skew-symmetric form on $\mathcal{M}$, apart from multiplication by a constant. Since any quantization of $\mathcal{M}$ along the general lines indicated in the last chapter (more specifically, as a so-called linear Bose-Einstein field), which is Lorentz-invariant, must involve such a form, this implies the essential uniqueness of the commutation rules for the Bose-Einstein quantization of a scalar field.

There is another way of formulating the representation space for the foregoing representation that is more familiar and has the advantage that the operation j on which the complex structure is based is represented simply by numerical multiplication by the complex number i. If we take a function ϕ in M, and restrict it to the positive-frequency part (i.e. $k_0 > 0$) of the mass-hyperboloid $k^2 = m^2$, we get a function ϕ' say satisfying the normalizability condition

$$\int_{k_0 > 0;\, k^2 = m^2} |\phi'(k)|^2 \frac{d_3k}{|k_0|} < \infty;$$

and conversely, every complex-valued function ϕ that is measurable and satisfies this condition, arises in this manner from a unique ϕ'. The set of all such ϕ' then forms a complex Hilbert space $\mathcal{M}'$ of the conventional sort, the inner product being defined in the obvious manner, by the equation:

$$(\phi', \psi') = \int \phi'(k)\psi'(k)^* \frac{d_3k}{|k_0|}.$$

These functions ϕ' are commonly called *scalar particle wave functions.* Their use in the present connection makes evident the positivity of the energy of the field, a simple but rather fundamental point. The energy of a system which is defined by equations invariant under translations in time, is commonly defined in quantum theory as the self-adjoint generator of the one-parameter group of unitary transformations representing translations in time, in their action on the wave functions of the system. There is no difficulty in checking that the energy as thus defined is represented by the following operator on M' : $\tilde{\phi}(k) \to k_0\tilde{\phi}(k)$. Obviously this operator is non-negative, having in fact a continuous spectrum consisting of the interval $[m, \infty)$.

Let us now consider the foregoing rigorous developments in relation to the heuristic notion of quantization. In its most elementary form the latter involves the replacement of the real-valued solution ϕ of the Klein-Gordon equation by an hermitian operator-valued solution, say $\boldsymbol{\phi}$, of the same equation. Now $\boldsymbol{\phi}$ will represent a classical mechanical system, and in fact essentially that represented by ϕ, unless some of the commutators $[\boldsymbol{\phi}(x), \boldsymbol{\phi}(x')]$ are distinct from zero. The simplest non-trivial assumption about these commutators is that they are numbers, so that

$$[\boldsymbol{\phi}(x), \boldsymbol{\phi}(x')] = iD(x, x'),$$

where $D(x, x')$ is real-valued (by virtue of the hermitian character of the $\boldsymbol{\phi}(x)$), and not identically zero. Now for such relations to be consistent from the viewpoint of Lie algebra it is necessary and sufficient, in a formal way, that the skew-symmetry and Jacobi conditions be satisfied. The former simply gives the restriction

$$D(x, x') = -D(x', x),$$

while the latter is satisfied automatically, by virtue of the commutativity of any operator with a number. Thus a formal quantization may be effected starting from any skew-symmetric function $D(x, x')$. But if it is required in addition that the commutation rules be Lorentz-invariant, i.e.

$$[\boldsymbol{\phi}(Tx), \boldsymbol{\phi}(Tx')] = [\boldsymbol{\phi}(x), \boldsymbol{\phi}(x')]$$

for any orthochronous Lorentz transformation T, then $D(x, x')$ must have the form

$$D(x, x') = \Delta(x - x'),$$

where Δ is a function of a single vector that is invarant under the orthochronous homogeneous Lorentz group. The uniqueness of the invariant bilinear skew-symmetric form on M that is Lorentz-invariant means essentially the uniqueness of the Δ of this type which is odd. The Δ used implicitly above is a generalized function (a distribution, actually, in this particular case) differing by a real constant multiple from the function of x given by the formula $\int_{k^2=m^2} e^{ix\cdot k}\, i\varepsilon(k)\, d_3k/|k_0|$.

To see this, we must consider how the purely formal quantized field $\phi(x)$ is connected with a canonical system of hermitian operators over the space $\mathscr{M}$, relative to the bilinear form B. This is important also to connect the conventional formalism in theoretical physics with the present rigorous formulation. Despite the fact that $\phi(x)$ has no known empirical physical or mathematical interpretation—in fact Bohr and Rosenfeld in a classic paper thirty years ago showed rather conclusively the former even in the more favorable case of the electromagnetic field, while the mathematical investigations of the past thirty years have uniformly indicated that $\phi(x)$ has no effective meaning (less so even than the dx in the expression dy/dx)—the notation is still extensively used in theoretical physics. Those using the notation will often agree that it is only a space-time average $\int \phi(x) f(x)\, d_4x$, where the weighting function f is, say, infinitely differentiable and vanishes outside a compact set, that has empirical and/or mathematical meaning; and claim on occasion that as long as one is aware of this, no harm can result from the use of the $\phi(x)$. However, it is obviously out of place in a logical theory, and its continued presence in theoretical work is a constant temptation to form trilinear products like $\phi(x)\psi(x)\psi(x)^*$, where ψ is another quantized field, which products, however extensive a role they play in the conventional treatments of quantum field-interactions, cannot be given empirical or mathematical meaning by any known process, including averaging with respect to a smooth weight function in a small region. Therefore we shall use the notion and notation $\phi(x)$ only in showing the connection between the conventional and rigorous approaches.

Let us assume now that we have obtained a mapping R from $\mathscr{M}$ to self-adjoint operators in some Hilbert space, such that R is real-linear (i.e. $R(f + g) \subseteq R(f) + R(g)$ and $R(af) \subseteq aR(f)$ if a is real), and the canonical commutation relations are valid:

$$R(f), R(g)] = -iB(f, g);$$

we may as well use the infinitesimal form of the relations involving R, since to trace the connection with the conventional formulation is necessarily a heuristic process. Since R is a linear function on the space $\mathcal{M}$, as a real Hilbert space, it is reasonable to write, in a figurative rather than mathematical sense,

$$R(\bar{f}) = \int_{k^2=m^2} \bar{f}(k)\bar{\Phi}(k)\,\frac{d_3k}{|k_0|},$$

where $\bar{\Phi}$ is a function whose values are hermitian operators. This function $\bar{\Phi}$ is then the conventional quantized field as a function on momentum space. To obtain the field in physical space, it is only necessary to take the Fourier transform:

$$\Phi(x) = \int_{k^2=m^2} e^{ix\cdot k}\bar{\Phi}(k)\,\frac{d_3k}{|k_0|}.$$

Furthermore, if g is any function on physical space having a continuous Fourier transform whose restriction to the mass hyperboloid is square-integrable (e.g., it suffices if g and all its partial derivatives up to third order, assumed to exist, are integrable over space-time), the figurative expression $\int\Phi(x)g(x)d_4x$ may be given a rigorous mathematical interpretation, as $R(\bar{h})$, where $\bar{h}^*$ is that element of $\mathcal{M}$ which coincides on the mass hyperboloid with the inverse Fourier transform of g; the validity of this interpretation follows at once from the figurative use of the Parseval formula. Another figurative use of the same formula yields readily the stated commutation relation between $\Phi(x)$ and $\Phi(x')$.

This discussion, although confined to the case of the Klein-Gordon equation, applies with little change to the case of any conventional relativistic particle and especially one of "integral spin," such as the photon. The quantization of the Maxwell equations may in particular be effected entirely analogously. We shall describe this and the electron wave field more because of the importance of these fields, rather than because any significant new problems arise (except for the change to Fermi-Dirac statistics in the case of the electron—which amounts, very roughly, to a change of sign, skew-symmetric forms being replaced by symmetric ones, etc.). The commutator function Δ in the general case of a system defined by a linear hyperbolic partial differential equation is replaced by a generalized function that is essentially the Riemann function for the equation. In the variable coefficient case one must deal with a function $D(x, x')$ of two vectors, which in the case of a second-order equation may be defined in quasi-heuristic fashion by the following Cauchy (initial value) problem: for each fixed x', $D(x, x')$ satisfies the equation as a function of x; $[D(x, x')]_{x_0=x_0'} = 0$; and $[\partial D(x, x')/\partial x_0]_{x_0=x_0'} = \delta(\mathbf{x} - \mathbf{x}')$, where $\mathbf{x} = (x_1, x_2, x_3)$. In addition to providing the commutator function for the corresponding quantized fields, this function can be used to solve the Cauchy problem for the equation with general initial data in a well-known fashion. In sufficiently simple cases this function can be shown to have rigorous existence as a distribution, but in general its singularities are too severe, and the precise global formulations applicable to general hyperbolic equations are still under development.

A conventional relativistic particle or field is describable by a system of constant-coefficients hyperbolic equations, which serves to pick out a subspace of the direct product of the space of all scalar functions on space time with a finite-dimensional representation space (or so-called "spin space") of the Lorentz group which is invariant under the group. Elementarity of the particle or field corresponds to the irreducibility of the representation of the Lorentz group defined thereby by the equation. The Maxwell and Dirac equations are the best known of such equations, and next to the Klein-Gordon equation just treated, are mathematically the simplest (apart from certain complications resulting from the vanishing mass of the photon). The finite-dimensional representations in question all admit invariant symmetric bilinear forms, which are necessarily indefinite, since the Lorentz group has no non-trivial finite-dimensional orthogonal representation. A remarkable and important feature of the relativistic equations is the existence of a positive definite symmetric bilinear form on the space of solutions, where a priori only an indefinite form was to be expected. In fact, the restriction of this indefinite form to the "real mass" wave function subspace, the Lorentz-invariant subspace in which all physically relevant fields may reasonably be expected to lie, turns out to be definite.

The Maxwell equations are somewhat atypical, but are of course a most important special case, and the general non-scalar representation of the Lorentz group can be treated along lines which are at any rate not more complicated. They may be described either in terms of potentials or in terms of field strengths; for the later discussion of quantum electrodynamics, the potentials must however be used. Consider then the space of all functions $A_k(x)$ on space-time ($k = 0, 1, 2, 3$; $x = (x_0, x_1, x_2, x_3)$) whose values are in the real vector representation space of the Lorentz group (which space is isomorphic to four-dimensional space time except that the translation subgroup of the Lorentz group acts only trivially). Under the action of the Lorentz group this space is far from irreducible, splitting as a direct integral, or continuous direct sum, of subspaces which derive essentially from an invariant quadratic expression in the generators of the

Lorentz group, and related expressions, which must act as constants in an irreducible subspace, the expression of this requirement being the partial differential equations defining the elementary vector particles. This approach could be rigorously developed, but for brevity we shall merely describe the Maxwell equations and the Lorentz-invariant forms on the space of its solutions, in terms of functions on "momentum space," i.e. the dual to space-time as a linear space.

Consider the class $\mathscr{C}$ of all complex-valued measurable functions $A_i(k)$ defined on the cone $C: k^2 = 0$ $(k = (k_0, k_1, k_2, k_3)$, $k^2 = k_0^2 - k_1^2 - k_2^2 - k_3^2)$, which satisfy the hermitian condition

$$A_i(-k) = A_i(k)^*,$$

and the linear condition

$$k \cdot A(k) = 0$$

(here the Lorentz-invariant scalar product is employed), almost everywhere on the cone. (That only functions on C are involved corresponds to the equations $\Box A_i = 0$ $(i = 1, 2, 3, 4)$ in physical space, while the linear condition expresses the side-condition $\sum_i \partial A_i/\partial x_i = 0$.) From the Cauchy-Schwarz inequality and the side condition it follows without difficulty that

$$A(k) \cdot A(k) \geqq 0.$$

The subset $\mathscr{L}$ of all elements A of $\mathscr{C}$ for which

$$\langle A, A \rangle = \int_C A(k) \cdot A(k)\, dk$$

is finite, where dk refers to the unique Lorentz-invariant element of volume on C, namely $|k_0|^{-1}\, dk_1 dk_2 dk_3$, is thereby well-defined. Setting $\mathscr{N}$ for the set of elements A of $\mathscr{L}$ such that $\langle A, A \rangle = 0$, $\mathscr{N}$ is actually a subspace of $\mathscr{L}$ and the quotient space $\mathscr{L}/\mathscr{N}$ is a real Hilbert space $\mathscr{H}$, the space of "normalizable photon fields," relative to the inner product for any two elements $\tilde{A}$ and $\tilde{A}'$ being defined as

$$\langle \tilde{A}, \tilde{A}' \rangle = \int_C A(k) \cdot A'(k)\, dk$$

where A and A' are any representatives for $\tilde{A}$ and $\tilde{A}'$ in $\mathscr{L}$. The ambiguity in the choice of A or A' is essentially the so-called "gauge-invariance of the second kind" in conventional theory.

As in the case of the Klein-Gordon equation, there is an essentially unique complex structure which may be imposed on the space, namely

$$j: \tilde{A} \to \tilde{A}', \quad \text{where } A'_j(k) = i\epsilon(k) A_j(k) \quad (j = 1, 2, 3, 4),$$

it being readily verified that j is well-defined modulo $\mathcal{N}$. Noting that $j^2 = -I$, that $\langle j\tilde{A}, j\tilde{A}'\rangle = \langle \tilde{A}, \tilde{A}'\rangle$, and that j commutes with the action of the proper Lorentz group, it then follows, in a familiar algebraic fashion, that $\mathscr{H}$ is a complex Hilbert space with respect to j as multiplication by i, and the inner product

$$(\tilde{A}, \tilde{A}') = \langle A, A'\rangle + i\langle jA, A'\rangle.$$

This has been a Lorentz-invariant construction, so that the action of the Lorentz group on $\mathscr{H}$ as a complex Hilbert space gives a unitary representation of the proper inhomogeneous Lorentz group. The elementarity of the photon corresponds to the essential irreducibility of this representation; more exactly, it splits into two irreducible subspaces, corresponding to different polarization states, which are interchanged by space reversal (i.e. the operation taking $x_0 \to x_0$ and $x_r \to -x_r$ for $r \neq 0$). The only mathematically difficult point here is the irreducibility, which again is a special case of the irreducibility established in Professor Mackey's chapters for a general class of representations of semi-direct products. To summarize:

THE MAXWELL REPRESENTATION: *The real normalizable solutions of the Maxwell equations in vacuum (with Lorentz side condition) form a complex Hilbert space on which the orthochronous Lorentz group acts in a continuous irreducible unitary fashion. Time reversal acts as a conjugation. The energy is a non-negative operator of continuous spectrum* $(0, \infty)$.

The assertions about time reversal and the energy have not been explicitly established above, but are simple corollaries. The positivity of the energy may be seen most readily by using a representation in which j acts as ordinary numerical multiplications by i, which is in fact the conventional positive-frequency representation for normalizable photon wave functions. The map $A_r(k) \to B_r(k) = (1/2)(A_r(k) + A_r(-k))$ carries $\mathscr{H}$ (the *real* photon fields) into a collection of complex-valued functions on the positive-frequency cone $C_+, k_0 > 0$ (the complex, positive-frequency fields), in such a fashion that the inner product (A, A') corresponds to the conventional type

$$(B, B') = \int_{C_+} B(k)\cdot B(k)\, dk,$$

and j corresponds to multiplication by i. That is, we have an ordinary Hilbert space (more precisely a subspace of such) of all complex-vector-valued functions on C_+, with a corresponding norm. Since the energy is represented by multiplication by k_0, the assertions above follow.

As regards time reversal, i.e. the transformation $x_0 \to -x_0$ and $x_r \to x_r$, this cannot be represented by a unitary operator, since it takes the energy into its negative, and a unitary representative for it would then set up a unitary equivalence between an operator with positive spectrum and one with negative spectrum. That the operation indicated is in fact a conjugation on the space $\mathscr{H}$ set up above of real normalizable photon fields is a matter of straightforward verification.

The Dirac equation, or relativistic wave equation for a so-called spin 1/2 particle, differs qualitatively from the Maxwell and Klein-Gordon equations most notably in the following features: it does not give a (single-valued) representation of the proper Lorentz group, but rather of its simply connected covering group; the corresponding energy is not non-negative. These differences are connected with the circumstance that Fermi-Dirac, rather than Bose-Einstein quantization, is appropriate for Dirac particles. Partly for brevity and partly because we are primarily concerned with the analytical rather than algebraic problems of quantum field theory, which are essentially statistics-independent, the Dirac equation will be dealt with only in general terms.

The relevant finite-dimensional representation, the so-called "spin" representation, is a rather extraordinary one not readily describable in intuitive terms. To arrive at this representation, let M be a real linear space of finite even dimension, with a distinguished real non-degenerate symmetric form (x, y); the associated Clifford algebra may be defined as the essentially unique one over the complex field generated by a unit and elements $f(x)$ ($x \in M$, f being a linear map on M) satisfying the relations

$$f(x)f(y) + f(y)f(x) = (x, y).$$

Adjunction on this algebra is the unique anti-involution leaving fixed the elements $f(x)$. The Clifford algebra is known to be isomorphic to the algebra of all linear transformations on a linear space N of dimension 2^m, where m is half the dimension of M, and on which there is a distinguished non-degenerate hermitian form, adjunction relative to which corresponds to the original operation in the Clifford algebra. A spinor is, in the first instance, an element of N, a spinor field being a function on space-time with values in N.

Now if O is any transformation on M leaving invariant the fundamental symmetric form, i.e. a "pseudo-orthogonal" transformation, the map $f(x) \to f(Ox)$ evidently extends to an automorphism of the Clifford algebra, and thereby a representation of the pseudo-orthogonal group on M by automorphisms of the Clifford algebra is set up. On the other hand, since every automorphism of a full matrix algebra is inner, there corresponds to O a linear transformation T_O on the spin space which induces the corresponding automorphism of the matrix algebra. Now T_O will only be unique within multiplication by a scalar, but since T_O may be chosen to be unitary relative to the form on N, this scalar may be limited to have absolute value unity. Corresponding to the automorphic representation described, there is then a scalar $c_{O,O'}$ such that

$$T_O T_{O'} = c_{O,O'} T_{OO'}.$$

A closer examination of the situation shows that with suitable choices for the T_O, the scalars $c_{O,O'}$ will always be ± 1; but that there is no choice for them which will result in a strict representation, with $c_{O,O'} = 1$. This explains the term "double-valued representation." By continuity, $c_{O,O'}$ will be unity if O and O' are sufficiently close to the identity, so that a local representation is obtained, which extends to the simply connected covering group of the pseudo-orthogonal group; this may be seen to cover it twice if M has dimension greater than two.

In the particular case when M is Minkowski space-time and the fundamental form the usual relativistic one, the pseudo-orthogonal group is the homogeneous Lorentz group, and there results a representation of the covering group on a four-dimensional spin space, on which there is defined a distinguished indefinite non-degenerate hermitian form. The Dirac equation describes an irreducibly invariant subspace of the direct product of this space with the space of all scalar functions on space time, under the action of the full group, including reversal operations. Specifically, choosing unit vectors e_r $(r = 0, 1, 2, 3)$ in the x_r directions, and setting $\gamma_r = f(e_r)$, so that $\gamma_r\gamma_s + \gamma_s\gamma_r = 2g_{rs}$, with $g_{rs} = 0$ for $r \neq s$ and $g_{00} = -1$, $g_{rr} = 1$ $(r > 0)$, let $\mathscr{C}$ be the collection of all measurable functions $\psi(k)$ on the "mass hyperboloid": $C_m: k^2 = m^2$, to the spin space described above, such that $\nabla\psi + im\psi = 0$, where $\nabla = \sum \gamma_r k_r$, for which

$$\|\psi\|^2 = \int_{C_m} \langle \psi(k), \psi(k) \rangle \, dk,$$

where $\langle \cdot, \cdot \rangle$ refers to the Lorentz-invariant hermitian form on the spin space and dk is the Lorentz-invariant element of measure on C_m, is finite (the indicated inner product being necessarily non-negative by virtue of the Dirac equation being satisfied). The action of the orthochronous Lorentz group on $\mathscr{C}$ is then unitary, continuous, and together with the anti-unitary action of time reversal, irreducible.

The energy spectrum is readily seen to run from m to ∞ and from $-\infty$ to $-m$. The apparent problem posed by these negative energies disappears with the appropriate Fermi-Dirac quantization, which gives a field with positive energy. On the other hand, if the Dirac particle is quantized relative to Bose-Einstein statistics, its energy is negative, which is in a way a deduction of the proper statistics for a Dirac particle from general physical principles, a result due originally to Pauli.

Notes to Chapter III

The question of the extent to which the partial differential equations for a relativistic elementary particle contain some of the theoretical physics of the particle not already present in its transformation law is an interesting one. There is no doubt that the representation alone (more precisely, the unitary equivalence class of the representation) suffices to determine the energy and momenta operators (as the infinitesimal generators of the group action in the representation), and that the representation is a unique attribute of a standard relativistic particle. There is also no doubt that there is no immediate prospect for empirical significance for the values of wave functions of these particles as functions on space-time, even after suitable averaging over a space-time region. Further, the quantum numbers, so-called, and the quantization rules, for a free field of such particles, are determined by the associated unitary representation of the Lorentz group. The theoretical formulation of a relativistic elementary free particle species as an irreducible unitary representation of the Lorentz group seems on the whole physically justified.

On the other hand, the conventional interactions between such particles cannot be readily expressed in terms of the corresponding unitary representations of the Lorentz group, in fact, the notion of a "local" interaction, such as have been conventionally assumed to be the only physical type, depends explicitly on the formulation of the vectors in the representation space as functions on space-time. The interaction is, however, not really a property of the free particles, and there is, moreover, no mathematically rigorous way at present to set up a relativistic interaction of the type described.

In any event, "free physical particle" is more of an analytical than a real concept, and it is helpful to have a clean-cut theoretical definition for the notion. At the same time, it needs to be borne in mind that such objects are really adjuncts to more complex structures, arising in all probability as the "quanta" of non-linear fields. The special feature of the quanta of the conventional fields is that a complete set of quantum numbers, i.e. generators of a maximal abelian algebra of observables on the single-particle space $\mathscr{H}$, can be defined which are all of group-theoretic origin.

References to Chapter III

While the connection of the standard relativistic equations with irreducible unitary representations of the Lorentz group has been known for more than two decades, there appears to be no detailed exposition of this material in book form from the group representation point of view. Following the realization of the role of group representation theory in atomic physics, and especially the success of the Dirac equation, a number of scientists, among them Dirac and von Neumann, appear to have become aware of the significance for relativistic quantum theory of the theory of unitary representations of the Lorentz group. Such representations were necessarily infinite-dimensional, in contrast to the familiar finite-dimensional non-unitary representations. A large number of representations of possible physical interest were formulated in the late 20s and early 30s, and a systematic determination of all unitary representations was undertaken by Wigner (1939), whose work was on a notably higher level of rigor than the bulk of the theoretical literature. The theory of Mackey was applicable to so-called semi-direct products of fairly general categories of groups, and covered in particular the Lorentz group, which is the semi-direct product of the homogeneous Lorentz group and the group of all translations in space-time. The results in particular substantially subsumed those of Wigner and at the same time provided a mathematically unexceptionable basis for them.

CHAPTER IV

General Structure of Bose-Einstein Fields

In the last chapter we went into some detail about the structure of the classical phase space associated with a neutral scalar relativistic particle, as well as about relativistic particles somewhat more generally. The quantization of the neutral scalar field was briefly indicated, but not at all completed, fundamental questions concerning existence and uniqueness of systems of canonical operators having been left unsettled. These questions have actually nothing at all to do with the specific character of the field being quantized, being identical for scalar mesons, photons, or fields defined on completely different spaces from conventional space-time and not necessarily defined by a partial differential equation. What matters mainly is that we are given a linear vector space $\mathcal{M}$, which may be thought of as the space of all classical fields of a certain variety, together with a distinguished skew-symmetric form, or some similar elements of structure, over $\mathcal{M}$. The present chapter aims to treat the quantization problem from this quite general point of view, the generality being useful not only in enabling one to treat a greater variety of fields, but also in clarifying the logical situation and in facilitating a compact mathematical treatment.

Let then $\mathcal{M}$ be a given real-linear vector space, and B a given non-degenerate real skew-symmetric form on $\mathcal{M}$. By "non-degenerate" is meant that the only vector z such that $B(z, z') = 0$ for all z' in $\mathcal{M}$ is $z = 0$. The reader anxious to keep his feet firmly on the ground may think of $\mathcal{M}$ as the space of all real normalizable scalar meson wave functions of a given mass and of B as the bilinear form given by the integral expression whose kernel is the usual commutator function. A *Bose-Einstein canonical system*, or for short and because of mathematical aspects, a *Weyl system*, over $(\mathcal{M}, B)$ is defined (as a purely mathematical object) as a map $z \to U(z)$ of $\mathcal{M}$ into the unitary operators on a complex Hilbert space $\mathcal{K}$, which is continuous in the weak operator topology as a function of z, when z is restricted to (each and every) finite-dimensional subspace of $\mathcal{M}$, and which satisfies the "Weyl relations":

$$U(z)U(z') = e^{(i/2)B(z,z')}U(z + z').$$

For any such system, it is immediate that $[U(tz); -\infty < t < \infty]$ is a continuous one-parameter unitary group, and so has a self-adjoint generator $R(z)$. These are the so-called field variables, i.e. the reader may think, if he wishes, of $R(z)$ as the average of the quantized scalar meson field with respect to a suitable weighting function (any whose Fourier transform agrees on the mass hyperboloid with that of z, to be specific).

It is natural a priori as well as relevant in the light of later developments to consider the existence, uniqueness, and transformation properties of Bose-Einstein canonical systems. The finite-dimensional case has already been treated somewhat; it is the infinite-dimensional case with which this chapter is primarily concerned.

To begin with the existence question, consider the situation for a finite-dimensional subspace $\mathcal{N}$ of $\mathcal{M}$ with the property that the restriction of B to $\mathcal{N}$ is non-degenerate. We know then that there exists a canonical system $R_{\mathcal{N}}(z)$, defined for all z in $\mathcal{N}$, and satisfying the Weyl relations. The set of all bounded functions of the $R_{\mathcal{N}}(z)$ forms a certain C^*-algebra which may be designated $\mathcal{A}_{\mathcal{N}}$. Now if $\mathcal{N}'$ is any finite-dimensional subspace containing $\mathcal{N}$, and on which B is likewise non-degenerate, the restriction of the canonical system $R_{\mathcal{N}'}(z)$ defined for all z in $\mathcal{N}'$ to the z in $\mathcal{N}$ will have the same general properties as the $R_{\mathcal{N}}(z)$. From the Stone-von Neumann uniqueness theorem it can be inferred that there must then exist an algebraic isomorphism from $\mathcal{A}_{\mathcal{N}}$ into $\mathcal{A}_{\mathcal{N}'}$ which essentially carries $R_{\mathcal{N}}(z)$ into $R_{\mathcal{N}'}(z)$, "essentially" referring to the circumstance that the $R(z)$ are unbounded, and so not really in the $\mathcal{A}$'s, what is really meant being that $f(R_{\mathcal{N}}(z))$ is carried into $f(R_{\mathcal{N}'}(z))$ for, say, all bounded continuous functions f.

Now it is not difficult to show that every vector z in $\mathcal{M}$ is contained in some subspace $\mathcal{N}$ on which B is non-degenerate and that, moreover, the collection of all such subspaces is directed: if $\mathcal{N}$ and $\mathcal{N}'$ are such, then there exists another such subspace $\mathcal{N}''$ which contains both $\mathcal{N}$ and $\mathcal{N}'$. Furthermore the isomorphism mentioned above of $\mathcal{A}_{\mathcal{N}}$ into $\mathcal{A}_{\mathcal{N}'}$ for the case $\mathcal{N} \subset \mathcal{N}'$ is unique as a consequence of the fact that an automorphism of any $\mathcal{A}_{\mathcal{N}}$ which leaves fixed (essentially) the $R_{\mathcal{N}}(z)$ must be the identity, as follows in turn from the existence of an irreducible representation for $\mathcal{A}_{\mathcal{N}}$. It results that, all in all, what we have is something like an ascending chain of subalgebras. More specifically, there is an isomorphism, say $\theta_{\mathcal{N}\mathcal{N}'}$, of $\mathcal{A}_{\mathcal{N}}$ into $\mathcal{A}_{\mathcal{N}'}$, when $\mathcal{N} \subset \mathcal{N}'$; if in addition, $\mathcal{N}' \subset \mathcal{N}''$, then

$$\theta_{\mathcal{N}'\mathcal{N}''}\theta_{\mathcal{N}\mathcal{N}'} = \theta_{\mathcal{N}\mathcal{N}''}.$$

That is to say, the various injections of the $\mathscr{A}_{\mathscr{N}}$ into the $\mathscr{A}_{\mathscr{N}'}$ match together.

This means that it is straightforward to form an algebraic "direct limit" $\mathscr{A}_0$ of the directed system of algebras-cum-injection-isomorphisms $[\mathscr{A}_{\mathscr{N}}, \theta_{\mathscr{N}\mathscr{N}'}]$; there is an injection of each $\mathscr{A}_{\mathscr{N}}$ into $\mathscr{A}_0$, these various injections match together with the $\theta_{\mathscr{N}\mathscr{N}'}$ in the obvious manner, and $\mathscr{A}_0$ is the set-theoretic union of the images of the $\mathscr{A}_{\mathscr{N}}$ under the injections. It is clear from its construction that $\mathscr{A}_0$ is an algebra with an involution $*$, the adjunction operation. Each element A of $\mathscr{A}_0$ has a norm, its bound as an element of some $\mathscr{A}_{\mathscr{N}}$ (which is independent of $\mathscr{N}$, providing only that A is actually an element of $\mathscr{A}_{\mathscr{N}}$), and these norms have the characteristic properties of the operator bound: $\|AA^*\| = \|A\| \, \|A^*\|$, etc. $\mathscr{A}_0$ will not be complete, but it may be completed with respect to the norm, yielding an algebra $\mathscr{A}$, and when this is done the properties of the norm are retained. It then follows from the Gelfand-Neumark abstract characterization of C^*-algebras (or alternatively by a direct argument about limits of concrete C^*-algebras) that $\mathscr{A}$ is a C^*-algebra. That is, there exists a Hilbert space $\mathscr{K}$ (not necessarily, and in fact very far from being, unique) such that the elements of $\mathscr{A}$ are isomorphically represented, with regard to algebraic operations and the norm, as bounded operators on $\mathscr{K}$.

Thus we obtain a map $z \to U(z)$ of $\mathscr{M}$ into the unitary operators on $\mathscr{K}$, satisfying the Weyl relations. However, there can be no a priori certainty of the validity of the continuity requirement for a Weyl system. While this might appear to be a minor matter, the fact is that the continuity will definitely not hold for every representation space $\mathscr{K}$, even when it holds for some $\mathscr{K}$; and in any event it is not clear whether a $\mathscr{K}$ always exists for which the continuity is satisfied. This question is equivalent to that of the existence of a *regular* state of $\mathscr{A}$, i.e. one whose restriction to each $\mathscr{A}_{\mathscr{M}}$ is regular in the sense previously indicated; all that the general theory guarantees is the existence of some state, quite possibly devoid of any regularity properties.

This is a more substantial m°'ter than appears at first glance. Some further assumption on the system $(\mathscr{M}, B)$ is apparently needed for the existence of a Weyl system. The assumption of the existence of a positive definite real symmetric form on $\mathscr{M}$, relative to which B is continuous, is one that is obviously satisfied in the concrete cases that are at all likely to come up, and is rather weak in a theoretical way. This relatively general case is however reducible, as regards the existence question, to the special case in which $\mathscr{M}$ admits the structure of a unitary (or not necessarily complete complex Hilbert) space, in such a fashion that B is given as the imaginary part of the inner product. For brevity, and because this is an extremely important special case for applications (in many of which there is a distinguished complex Hilbert space structure that is physically significant), we discuss only this case, and state specifically

EXISTENCE THEOREM. *For any complex Hilbert space $\mathscr{H}$, there exists a Weyl system over $(\mathscr{H}, B)$, where $B(z, z') = \operatorname{Im} \langle z, z' \rangle$, $\langle z, z' \rangle$ denoting the inner product in $\mathscr{H}$.*

Of course, the restriction of a Weyl system to a submanifold is again a Weyl system, so that the case of an incomplete unitary space follows automatically from the complete case covered by the theorem as stated. Any proof depends upon the implicit or explicit construction of a regular state for the algebra previously described, and will establish rather more than the theorem. In fact, one might as well set up the free-field representation at the same time that the theorem is proved. This representation will be discussed at length in a later chapter, and it should suffice here to indicate briefly how it may be obtained.

For a finite-dimensional subspace $\mathscr{N}$ of $\mathscr{M}$, there is a unique pure state of $\mathscr{A}_{\mathscr{N}}$ that is regular in the sense indicated earlier and invariant under all unitary transformations of $\mathscr{N}$ onto itself (or rather, under the automorphisms of $\mathscr{A}_{\mathscr{N}}$ induced by these unitary transformations). The uniqueness of this state $E_{\mathscr{N}}$ means that the various $E_{\mathscr{N}}$ match together; if $\mathscr{N}'$ is any other finite-dimensional complex-linear submanifold of $\mathscr{H}$, and if $\mathscr{N}''$ is another such manifold containing both $\mathscr{N}$ and $\mathscr{N}'$, then the restriction of $E_{\mathscr{N}''}$ to $\mathscr{A}_{\mathscr{N}}$ (or to $\mathscr{A}_{\mathscr{N}'}$) necessarily agrees with $E_{\mathscr{N}}$ (or $E_{\mathscr{N}'}$), by virtue of the stated group-theoretic characterization of these states. In this way a unique positive linear functional E_0 on $\mathscr{A}_0$ is obtained which essentially extends all the $E_{\mathscr{N}}$ and extends easily to a state E of $\mathscr{A}$, which is obviously regular by construction. The representation of $\mathscr{A}$ associated with E, by virtue of the general mutual correspondence between states and representations of C^*-algebras, will then give a Weyl system. It may be noted in passing that without a positive definite symmetric form on $\mathscr{M}$ (provided in this case by the real part of the inner product), no unique way of picking out states $E_{\mathscr{N}}$, i.e. of forming a coherent family of such states, is apparent.

Now let us turn to the uniqueness question. As indicated in the second chapter, uniqueness can not be expected in the conventional form. If multiplicity is suppressed by requiring irreducibility, this will not help; the examples given in the second chapter can be shown to be irreducible, about as closely connected as representations are likely to be, and yet are unitarily inequivalent. The solution to this difficulty is however conceptually so simple that, looking backwards, it seems that it should in essence have been discoverable at the outset of work on the problem, although there is no doubt that the felicitous development of the theory of C^*-algebras has made these matters seem very much simpler than they did at that time. The intuitive idea behind it is that the basic conceptually physically measurable quantities are the canonical field variables themselves; that in addition to these the smooth bounded functions of *finite* sets of them should be measurable (but not necessarily such functions of *infinite* sets); and that if a sequence of bounded observables converges uniformly, then the limit is an admissible observable, since its expectation value in any state can be determined to within an arbitrary degree of accuracy through the measurement of the approximating observables. Thus, more specifically, we first take all *bounded* functions of *finite* sets of field variables as the primary observables; and then, essentially as a matter of mathematical convenience, admit also the secondary observables obtainable as limits of uniformly convergent sequences of primary observables. This is a priori a physically natural procedure, on a mathematically naïve level; and it is reassuring that its technical implementation, while somewhat sophisticated, does indeed turn out to circumvent and illuminate the cited uniqueness difficulties.

The procedure sketched is similar to that initially investigated in connection with the existence question. As noted, its relative mathematical effectiveness is due to the availability of the theory of C^*- (i.e., uniformly closed, self-adjoint) algebras. While the difference between such an algebra and a ring of operators in the sense of Murray and von Neumann is in the beginning merely a matter of topology, there are great differences in their qualitative mathematical properties, which underly the relevance of the first as against the second in the present situation. The crucial difference is roughly that weak approximation of operators, in contrast to uniform approximation, has no direct physical significance for the corresponding observables—weak approximation depends in fact on the particular representation of the canonical variables, and moreover is affected by enlargement of the physical system under consideration, while uniform approximation is both independent of the particular representation of the canonical variables, and unaffected by enlargement of the physical system.

To proceed in a rigorous manner, we define a *field observable*, relative to a given Weyl system, as a bounded function of any finite set of canonical variables or as a limit of a uniformly convergent sequence of such. The collection of all such operators may be called the algebra of field observables, relative to the system. We can now state

UNIQUENESS THEOREM. *For any two Weyl systems R and R' over a classical linear phase space $(\mathcal{M}, B)$, there exists a unique algebraic isomorphism between the corresponding algebras of field observables that essentially carries $R(z)$ into $R'(z)$, for all elements z of M.*

We shall not go into the proof which, while not long, is perhaps slightly sophisticated. What this result means is that there is a *unique* abstract C^*-algebra of field observables, say $\mathscr{A}$, associated with a given classical linear system $(\mathcal{M}, B)$ (provided there exists any Weyl system at all over $(\mathcal{M}, B)$); this may be called THE algebra of field observables, or for short THE Weyl algebra, over $(\mathcal{M}, B)$. (A pure mathematician naturally asks whether a direct purely algebraic characterization of the Weyl algebra can be given, which materially avoids the use of concrete representations in Hilbert space; the answer is presently not clear, except that there are significant difficulties in the way of a purely algebraic, essentially different, characterization.) For the treatment of time reversal and other operations usually represented by anti-unitary transformations, it is useful to have a variation of the uniqueness theorem: if R and R' are as above except that R' is associated with the commutator form $-B$, then the same result holds except that in place of a straight algebraic isomorphism there is a conjugate-linear ring isomorphism.

It is interesting to note that the situation here is rather simpler in a way than the corresponding one in classical mechanics. If the form B is degenerate, substantially nothing can be said about uniqueness as is easily seen by examples. Now the more degenerate B is, the more classical is the system, the vanishing of B corresponding to a fully classical system. Thus the fully quantum-mechanical system is the one admitting the simplest relevant mathematical result.

To recapitulate, we have arrived at a unique way of "quantizing" a given classical linear field, which is satisfactory from a general phenomenological point of view. It remains now to examine the kinematical, statistical, and dynamical implications of this mode of quantization.

Taking first the kinematical situation, it is a formal commonplace in quantum field theory treatments that any unitary transformation on the classical (or "single-particle") space $\mathcal{M}$ (taking for the moment the case when $\mathcal{M}$ is a Hilbert space) induces a corresponding transformation on the field state vector space. This is frequently established in the crucial cases, such as translation in space-time, by actually writing down the rather ponderous, figurative expressions, unfortunately devoid of mathematical meaning or apparent prospect of such, obtained by substitution of the quantized for the classical field in classical expressions for the energy-momentum, etc. Actually the transformation of the field induced by a classical motion can be treated in a simple rigorous fashion through the use of the foregoing uniqueness result.

Let T be an invertible linear transformation on $\mathcal{M}$ which preserves the form B (a so-called "symplectic" transformation) and let f be an arbitrary linear functional on $\mathcal{M}$. If R is any Weyl system over $(\mathcal{M}, B)$, then defining R' by the equation

$$R'(z) = R(Tz) + f(z)I \qquad (z \in \mathcal{M}),$$

it is clear that R' is also a Weyl system (here I denotes the identity operator). Using the briefer term "motion" for what we previously called a "physical automorphism," it follows now from the uniqueness theorem that

COROLLARY ON MOTIONS OF THE WEYL ALGEBRA. *For any symplectic transformation T on the classical phase space $\mathcal{M}$ (relative to the non-degenerate skew form B on $\mathcal{M}$) and linear functional f, there exists a unique motion of the Weyl algebra that essentially carries $R(z)$ into $R(Tz) + f(z)I$, for all z in M.*

In this formulation, R refers to a perfectly arbitrary canonical system over $\mathcal{M}$; i.e. since we are dealing with the *abstract* algebra of field observables, the stated result holds for each and every concrete canonical system R over $(\mathcal{M}, B)$.

What this result means is that, although there will in general be no operator $\Gamma(T)$ acting on the state vectors (in a particular representation) that corresponds to, say, a symplectic transformation T on the classical space M, nevertheless the figurative expression

$$\Gamma(T)X\Gamma(T)^{-1}$$

will have a rigorous interpretation for every field observable X. Thus one has a well-defined motion of the observables, despite the lack of any but a figurative motion of the state vectors. It should be noted that the motion of the *states*, as opposed to *state vectors*, is well-defined also; the states merely transform contragrediently to the observables.

For the primary kinematical desideratum of relativistic field theory, that of showing that the Lorentz group acts as a group of motions of the field observables, the foregoing corollary to the uniqueness is rather more than adequate. It applies equally well to transformations unrelated to the action of the Lorentz group, as well as to physical models based on entirely different models of space-time and corresponding symmetry groups from the conventional Minkowski space—Lorentz group geometry. Any such kinematical transformation S will act as a homogeneous symplectic transformation on M and if $\gamma(S)$ is the automorphism corresponding to S^{-1} in the fashion indicated above, then $\gamma(SS') = \gamma(S)\gamma(S')$, i.e. the fundamental classical symmetry group acts as a group of motions on the field observables. One has, so to speak, a rigorous "automorphic" representation of the Lorentz group, on a well-defined algebra of observables, in place of a formally unitary representation on a rather vague Hilbert space of state vectors, in the case of the standard relativistic theories. (To avoid possible confusion, it may be useful to note here, although logically it is relevant later, that there will be representations in which the kinematical action of the Lorentz group will be unitary, for all the conventional linear fields; but these are not representations in which the temporal development for an interacting field of particles whose free behavior are described by the given linear fields can be unitary, or vice versa.)

While the statistics appear at first glance to be an entirely different matter from the kinematics, it may be partially subsumed under the general kinematics as formulated above, by taking the symplectic transformation T not as a Lorentz transformation (or more precisely, the action of such on $\mathcal{M}$), but as a type of "phase" transformation. The "occupation numbers" and particle interpretation are however definable only relative to a particular state of the system, the state of physical relevance being the so-called "physical vacuum" state. It therefore seems appropriate to postpone the treatment of statistics until after the consideration of the general role of the physical vacuum in the next chapter.

The dynamics of a Bose-Einstein field is a generally far less clear-cut question than any of those treated above. There are, however, a few simple and conservative statements that may be made about the qualitative character of the dynamics.

In the non-relativistic case, the temporal development of the field will be given as a one-parameter family of motions of the algebra of field observables. For the crucial case of a temporally homogeneous, or mechanically conservative, system, this family will in fact be a group. This group of motions is quite distinct from the groups of motions arising from the physical kinematics. The mode of establishment of this group in the special cases that have been of interest in quantum field theory is a separate type of question from that of the basic theory of the Weyl algebra. It will suffice here to remark that certain simple but typically divergent such motions can be rigorously formulated as one-parameter groups of inhomogeneous symplectic motions of a Weyl algebra.

In the relativistic case, the entire Lorentz group will act on the field observables by motions; this dynamical representation of the group must of course be distinguished from the kinematical representation described above. The kinematical (automorphic) representation, say $a \to \theta_0(a)$, where a is an arbitrary element of the Lorentz group and $\theta_0(a)$ is the corresponding kinematical motion of the algebra of field observables, is comparatively extremely accessible, and is describable in fairly explicit and entirely convergent analytical form. The dynamical representation, on the other hand, say $a \to \theta(a)$, is not at all determined by $(\mathcal{M}, B)$, but by the interaction; no non-trivial relativistic dynamical motion has yet been explicitly formulated in a fully satisfactory way. We shall have more to say about this later, but even in a general way a non-trivial restriction may readily be formulated for the dynamical representation.

This is the so-called "covariance" condition. From a conventional theoretical physical point of view this may be stated most succinctly, perhaps, as the condition that the Lagrangian should be an absolute invariant under the action of the Lorentz group. The Lagrangian, however, is just the sort of object that it seems desirable to avoid in a treatment aiming at mathematical rigor or clarity and empirical relevance or conceptual physical meaning. It is useful, therefore, to formulate the covariance in terms of the dynamical representation, which is mathematically a familiar kind of object and is close to what is measured, as quantum-theoretic concepts go. The condition is, in fact, quite simple in these terms:

$$\theta(aba^{-1}) = \theta_0(a)\theta(b)\theta_0(a)^{-1},$$

for arbitrary Lorentz transformations a and b. The limiting case of this when b is a displacement through an arbitrarily long time can be checked relatively directly, in a number of cases, amounting to the independence of transition probabilities from the Lorentz frame. More specifically, when such limits exist, the condition is equivalent to independence of the so-called forward and backward wave automorphisms ω_- and ω_+ from the Lorentz frame, from which follows the absolute Lorentz invariance of the scattering automorphism σ, where

$$\omega_\pm = \lim_{t \to \pm\infty} \theta(-t')\theta_0(t'),$$

$$\sigma = \omega_+^{-1}\omega_-$$

(t' denoting the Lorentz transformation consisting of translation through the time t); but the condition is quite independent of the possible existence of such limits.

Relative to a given dynamical representation $a \to \theta(a)$, the physical vacuum is definable as a regular state that is invariant under all of the $\theta(a)$ and for which the induced energy spectrum is non-negative. It is difficult not only to formulate the relevant dynamical representations, but also to establish the existence and/or uniqueness of the corresponding physical vacuums. Before discussing such matters further, the more approachable and equally inevitable matters of the general role of the physical vacuum and the structure of the free field will be explored.

Notes to Chapter IV

1. *Pathological canonical systems.* It is easy to give examples of self-adjoint operators P and Q having a common dense domain which they leave invariant, and whose commutator $PQ - QP$ agrees on this domain with iI, but which nevertheless do not satisfy the Weyl relations. For example, let $\mathscr{H}$ be the Hilbert space of all square-integrable functions on $(0, 1)$, and let $\mathscr{D}$ be the domain of all C^∞ functions vanishing near the endpoints. Let Q denote the (bounded) operator $f(x) \to xf(x)$ on $\mathscr{H}$, while P is to denote any self-adjoint extension of the operation which maps $\mathscr{D}$ as follows: $f(x) \to -if'(x)$.

On the other hand, one may readily set up one-parameter unitary groups $U(s)$ and $V(t)$ on a Hilbert space which satisfy the Weyl relations, but not the associated continuity conditions, so that the Heisenberg relation is inapplicable. E.g., let $\mathscr{H}$ denote the Hilbert space of mean-periodic functions of order 2 on the real line, and let $U(s)$ and $V(t)$ act as follows:

$$U(s), f(x) \to f(x+s); \qquad V(t), f(x) \to e^{itx}f(x).$$

2. ***Weak vs. uniform topology, and non-observable operators.*** Readers who have not had much occasion to become familiar with the weak and uniform operator topologies and their relations to algebras and to such results as the Stone-von Neumann uniqueness theorem for the Schrödinger operators, may find the technical situation somewhat elusive in a general way, in the present chapter. There is no real substitute for the cited mathematical background, but a few special explanatory notes may be useful.

The definition above of the Weyl algebra depends on the weak topology in the formulation of the $\mathscr{A}_{\mathscr{M}}$ for $\mathscr{M}$ finite-dimensional, a feature which at first glance might seem out of place in a construction which emphasizes the uniform topology. The point is that for finite-dimensional $\mathscr{M}$ there is no essential difference between the result obtained with the two topologies, for reasons given in the first chapter. A construction might have been employed which was based entirely on the uniform topology, and an algebra obtained for which results similar to those described above for the Weyl algebra, would be valid. However, the circumstance that the Weyl algebra includes all bounded functions of finite sets of canonical variables that could be even remotely relevant makes it more convenient. It ought perhaps to be noted again that the "field observables," such as canonical variables, while conceptually observable, are not the objects really measured empirically, whose theoretical counterparts are more complex, being associated with automorphisms of the algebra of field observables (or more precisely, the unitary operators representing these automorphisms in the representation determined by the physical vacuum state, as treated in the next chapter). These unitary operators, which give the real physical temporal development, would be the same whether the present Weyl algebra is used or any of various technical variants, some of which can be formulated entirely without the use of the weak topology.

Naïvely, one might also well ask whether the Weyl algebra is actually materially smaller than the algebra of all bounded operators, in, say, the "free-field" representation (which will be treated in detail in Chapter 6). This can be shown to be the case in several ways; in particular it can be shown that no non-trivial bounded function of the "total number of particles" (see below) is in the Weyl algebra. This has the simple if quite rough and somewhat oversimplified interpretation that the total number of "bare" particles is devoid of physical meaning.

References to Chapter IV

See Segal's paper (1959b) for a more detailed account of the basic material of this chapter, and Shale (1962) for a study of the automorphisms induced by various classes of symplectic transformations and their unitary implementability, etc.

CHAPTER V

The Clothed Linear Field

In the last chapter we treated the general phenomenology, kinematics, and dynamics of a Bose-Einstein field. Associated with any linear space $\mathscr{M}$ (representing all classical fields of a particular species) with a distinguished suitable bilinear form B (which is determined by the partial differential equation defining $\mathscr{M}$, in practice, and is physically analogous to the so-called fundamental bilinear covariant in classical mechanics, as well as related to the quantum-mechanical commutator), there was a unique algebra $\mathscr{A}$ of "field observables," consisting of all bounded functions of finite sets of the field variables $R(z)$, together with their limits in the sense of uniform convergence, satisfying the associated Weyl relations (i.e. field commutation relations in bounded covariant form). Any group of linear transformations on $\mathscr{M}$ leaving invariant B is then canonically represented by a group of automorphisms of $\mathscr{A}$, which gives the kinematics. The dynamics will, in nontrivial covariant cases, be given by a different group of automorphisms of $\mathscr{A}$, which must be given separately.

This is fine as far as it goes, but to make contact with empirical physics it is necessary to deal further with the determination of transition probabilities, the possible values of the energy and preferably to give a particle interpretation for the states of the quantum field, many experimental physicists considering the particles, rather than the field, to be fundamental. This may seem like a complicated task, but in essence it reduces to the treatment of the so-called physical vacuum. There are various ways of giving a theoretical definition for the vacuum, which is a certain distinguished state of $\mathscr{A}$, e.g. the presumably (or, at least, hopefully) "positive-energy," "regular" state invariant under (dynamical) translations in space-time, in the case of a Lorentz-invariant system; or, the state of lowest energy, in the conventional treatments (unfortunately, this is from a literal viewpoint mere rhetoric since in the interesting cases the energy is given by a formula which is mathematically meaningless); etc. In many ways, however, the concept of physical vacuum seems even more fundamental than that of field energy and it may very well be more sensible to take it as a primitive notion and deal with the energy, etc., in terms of it, rather than vice versa. In the interests of mathematical as well as physical conservatism, we shall proceed along these lines.

It is appropriate to begin with those properties of the vacuum which result simply from its being a state. Having considered these, the relation between the vacuum and a given temporal development for the field will be considered. More general transformation and statistical aspects of the physical vacuum will then be taken up. Finally, some examples of vacuum states and associated structures are developed, including a familiar so-called "divergent" field, whose treatment along the present lines is straightforward and rigorous.

Now let $\mathscr{A}$ be the Weyl algebra over the unitary space $\mathscr{H}$, and let E be a given physical vacuum state for $\mathscr{A}$. Then as noted in the first chapter, there is an associated representation for $\mathscr{A}$, simply by virtue of its being a C^*-algebra. To sketch briefly how this arises, we forget for the moment about $\mathscr{H}$, and form a new unitary space, consisting essentially of $\mathscr{A}$ itself, with the inner product

$$\langle A, B\rangle = E(B^*A);$$

all of the usual requirements on a unitary space are easily seen to be verified, except that $\langle A, A\rangle$ may vanish without A necessarily vanishing. If elements B and B' of $\mathscr{A}$ which differ by such a null vector A, $B' = B + A$, are identified, a unitary space in the strict sense is obtained. This space, say $\mathscr{K}_0$, is not necessarily complete but can be completed to a Hilbert space $\mathscr{K}$. There is an obvious "natural" map from $\mathscr{A}$ onto $\mathscr{K}_0$, say $A \to \eta(A)$ ($\eta(A)$ being the set of all elements of $\mathscr{A}$ equivalent to A). A representation of $\mathscr{A}$ is now obtained by making A correspond to the operator $\phi_0(A)$ on $\mathscr{K}_0$ defined as follows:

$$\phi_0(A)\colon \eta(B) \to \eta(AB).$$

It can be verified that $\phi_0(A)$ is in fact well-defined thereby, and that moreover ϕ_0 has the usual properties of a representation:

$$\phi_0(A + B) = \phi_0(A) + \phi_0(B), \qquad \phi_0(\alpha A) = \alpha\phi_0(A) \quad (\alpha \text{ a complex number}),$$
$$\phi_0(AB) = \phi_0(A)\phi_0(B), \qquad \phi_0(A^*) = (\phi_0(A))^*.$$

Furthermore, it can be shown that $\phi_0(A)$ is a bounded operator on $\mathscr{K}_0$, and so extends uniquely to a bounded operator on $\mathscr{K}$, which may be designated $\phi(A)$; ϕ is then a representation of $\mathscr{A}$ on a Hilbert space $\mathscr{K}$.

In addition to the representation ϕ, and the linear map on $\mathscr{A}$, there is an important distinguished vector v in $\mathscr{H}$, namely $\eta(I)$. v has the property that

$$E(A) = (\phi(A)v, v),$$

so that in the representation obtained (although generally not in an arbitrary representation), E can be represented by a normalizable vector in more-or-less the conventional quantum-mechanical manner. Furthermore, v is a "cyclic" vector for ϕ (which means, as a matter of definition, that the $\phi(A)v$ span $\mathscr{H}$). These properties can be shown to be characteristic. That is to say, given a representation ϕ' of $\mathscr{A}$ on a Hilbert space $\mathscr{H}'$ with a cyclic vector v' such that

$$E(A) = (\phi'(A)v, v),$$

the structure $(\mathscr{H}', \phi', v')$ is unitarily equivalent to the one constructed above. It is also worth noting, for background purposes, that the representation ϕ can be shown to be irreducible if and only if the state E is pure.

Now returning to the specific case in which $\mathscr{A}$ is the Weyl algebra over $\mathscr{H}$, being generated by the unitary operators $W(z)$ satisfying the Weyl relations

$$W(z)W(z') = \exp((i/2)\operatorname{Im}(z, z'))W(z + z'),$$

the unitary operators $W'(z) = \phi(W(z))$ will also satisfy the Weyl relations, and so define a Weyl system over $\mathscr{H}$ if and only if the continuity requirement is satisfied. This requirement, that $\{W'(tz); -\infty < t < \infty\}$ be for any z a weakly continuous function of t at $t = 0$, is however not automatically satisfied, and examples can be given to show that even in the case of a finite-dimensional Hilbert space, it may well be violated.

However, it turns out, quite conveniently from a technical standpoint, that reasonable physical desiderata on the state E, as the vacuum state of a physical system, are almost equivalent and in any event imply this continuity condition. In the first place, the stated continuity implies the continuity as a function of t, near $t = 0$ of $\exp[itR(z)]$, $R(z)$ being the corresponding self-adjoint field variable. Now $e^{itR(z)}$ depends on the "field" $R(z)$ and the parameter t in an almost maximally smooth and bounded way and it seems reasonable that its vacuum expectation value should have some corresponding such dependence. The stated condition seems maximally weak among simply formulable non-trivial such conditions. Thus, if E is an honest physical vacuum, the representation determined by it may quite reasonably be expected to determine a concrete Weyl system W' as the image under ϕ of the abstract system W.

To approach the continuity condition from another direction, if $\mathcal{M}$ is a finite-dimensional subspace of $\mathcal{H}$, then $\mathcal{A}_{\mathcal{M}}$ is a system of a finite number of degrees of freedom of the sort considered in Chapter I. The state E, when restricted to $\mathcal{A}_{\mathcal{M}}$ is a certain state $E_{\mathcal{M}}$ of this finite system. Now as indicated in Chapter I, there is good reason in an exact theory to require that any such state admit a certain elementary regularity property. Specifically, $E_{\mathcal{M}}$ should have the form

$$E_{\mathcal{M}}(A) = \operatorname{tr}(AD_{\mathcal{M}}),$$

$D_{\mathcal{M}}$ being an operator of absolutely convergent trace (relative to the ring $\mathcal{A}_{\mathcal{M}}$, i.e. in a representation for $\mathcal{A}_{\mathcal{M}}$ as the ring of all bounded operators on a Hilbert space). But if $E_{\mathcal{M}}$ has this form, then $E(W(tz)) = \operatorname{tr}(W(tz)D_{\mathcal{M}})$, taking $\mathcal{M}$ as the one-dimensional subspace spanned by z, for some operator $D_{\mathcal{M}}$ of absolutely convergent trace, and the continuity as a function of t follows.

In such a fashion one may be led to make the

Definition. A *regular state* E of a Weyl algebra $\mathcal{A}$ over a space $(\mathcal{H}, B)$ is one whose restriction $E_{\mathcal{M}}$ to the subalgebra $\mathcal{A}_{\mathcal{M}}$ relative to any finite-dimensional subspace on which B is non-degenerate has the form

$$E_{\mathcal{M}}(A) = \operatorname{tr}(AD_{\mathcal{M}}),$$

for some operator $D_{\mathcal{M}}$ of absolutely convergent trace relative to $\mathcal{A}_{\mathcal{M}}$.

This is mathematically a rather weak regularity condition, and it is surely reasonable from a physical viewpoint to expect a physical vacuum state to be regular in this sense. In a theoretical way the definition of regularity is somewhat further validated by its mathematical development. It will suffice for our purposes to cite the following results.

Theorem. *A state E of a Weyl algebra $\mathcal{A}$ over a space $(\mathcal{H}, B)$ is regular if and only if there is a concrete Weyl system W over $(\mathcal{H}, B)$ and a vector v in the representation space $\mathcal{K}$ for W, such that*

$$E(A) = (\phi(A)v, v),$$

where $\phi(A)$ is the concrete operator on $\mathcal{K}$ corresponding to the element A of the abstract Weyl algebra $\mathcal{A}$.

Alternatively, the state E is regular if and only if its generating functional $\mu(z) = E(W(z))$ is continuous relative to every finite-dimensional subspace of $\mathcal{H}$, and if in addition E is the "natural" state whose generating functional is $\mu(z)$.

More specifically, let a functional μ on $\mathscr{H}$ be called quasi-positive-definite in case it has the continuity feature just described, and if in addition, for arbitrary complex numbers ξ_j and vectors z_j in $\mathscr{H}$ $(j = 1, 2, \cdots, n)$,

$$\sum_{j,k} \mu(z_j - z_k)e^{iB(z_j, z_k)}\xi_j\bar{\xi}_k \geqq 0.$$

It is easily seen that the generating functional of a regular state must be quasi-positive-definite. Conversely, any such function determines in a "natural" way a regular state E of which it is the generating functional.

There is a mutual correspondence between regular states of the Weyl algebra and Weyl systems and there are a variety of reasons for postulating that the physical vacuum is regular. But regularity is only a mild smoothness condition on the vacuum and we have yet to examine the determining features of the vacuum. To this end we again forget about the space $\mathscr{H}$ and deal with $\mathscr{A}$ as an abstract C^*-algebra. A vacuum is, as a mathematical and physical object, defined relative to a particular motion of $\mathscr{A}$, which is given in non-relativistic form by a one-parameter group ζ_t of automorphisms of $\mathscr{A}$. The conventional formulation of the vacuum as "the state of lowest energy" is not effective as long as the energy of a state is not defined, and in fact "state" as used in this conventional formulation is not really the same as "state" as used here, but corresponds rather more to our notion "state vector," i.e. is a notion defined relative to a particular representation.

To arrive at a simple rigorous characterization of the vacuum consider first that a vacuum E must in any event be a stationary state, i.e. it must satisfy

$$E(\zeta_t(A)) = E(A)$$

for all t. Now for any such state the representation structure associated with E as above may be augmented by a one-parameter unitary group U_t $(-\infty < t < \infty)$ determined in the following way

$$\phi(\zeta_t(A))v = U_t\phi(A)v,$$

which implies in particular that U_t leaves v invariant, for all t. This group U_t may be constructed along lines generally similar to those briefly indicated. It will not necessarily be continuous, but if there is to be an energy operator, a rather modest desideratum, it must indeed be continuous, for the energy is the generator of this group U_t giving the temporal development of the system in the representation determined by the vacuum. This continuity is equivalent to the assumption, which we now make, that $E(\zeta_t(A)B)$ is for any A and B in $\mathscr{A}$ a continuous function of t.

The (clothed) hamiltonian or energy operator of the system is then the self-adjoint operator H given by Stone's theorem, such that $e^{itH} = U_t$, $-\infty < t < \infty$. Since the vacuum state representative, v, is invariant under the U_t, H annihilates v. The conventional formulation of the vacuum may now be utilized and E defined as a vacuum (relative to the given motion) in case v is the lowest "eigenstate" of H, or in other terms, if H is a non-negative self-adjoint operator.

Both mathematically and physically it is clear that in general a vacuum state neither exists nor is unique. Either feature represents a non-trivial property of the motion ζ_t. However, in relativistic quantum field theory, it may reasonably be anticipated, again for both sorts of reasons, that the vacuum will exist and be unique. On the other hand, there are other physical desiderata pertinent to the physical vacuum and these should be examined before going further with the given definition.

In the case of a relativistic system, there will be associated not only a one-parameter group $\{\zeta_t\}$ of automorphisms representing the temporal development of the system, but more inclusively, a representation $g \to \zeta_g$ of the entire Lorentz group by such automorphisms, agreeing on the time-translation subgroup with a non-relativistic one-parameter group of the sort just considered. It is plausible that the vacuum should be invariant under the full Lorentz group, i.e. that E should satisfy the equation

$$E(\zeta_g(A)) = E(A)$$

for all A in $\mathscr{A}$ and Lorentz transformations g. A priori one might expect to characterize the vacuum for a relativistic system through this requirement, and such a purely algebraic condition would appear to have some advantages over the one previously given, which can not be stated in such direct terms involving as it does the representation structure associated with the state. However, it turns out that although no Lorentz-invariant states other than the vacuum are known conventionally such do nevertheless exist. Thus, Lorentz-invariance is in itself insufficient to characterize the vacuum; positivity of the energy, as defined above, must be added and, when this is done, invariance under the temporal development is as effective as Lorentz-invariance.

A different sort of desideratum is the possibility of a particle interpretation. For making the connection with empirical physics, the particle interpretation is crucial. Nevertheless, its theoretical development and interaction with field theory in general has been rather difficult and quite tentative. It will be enough here to indicate what can be done with the general definition of occupation number given above and, for this purpose, let us assume, that as in conventional practice $(\mathscr{H}, B)$ is derived from a unitary space $\mathscr{H}$.

If $\mathscr{M}$ is any subspace of $\mathscr{H}$ of finite dimension or co-dimension, the operation $V_t: x \rightarrow e^{itP}x$, where P is the projection on $\mathscr{H}$ whose range is $\mathscr{M}$, is symplectic, and induces an automorphism ν_t of the Weyl algebra over $\mathscr{H}$. Occupation numbers relative to a physical vacuum E may now be defined in a heuristic way as follows. Let $[N_t; -\infty < t < \infty]$ be the one-parameter group on the representation space $\mathscr{K}$ associated with E as above, defined by the equation

$$N_t\phi(A)v = \phi(\nu_t(A))v,$$

A being arbitrary in $\mathscr{A}$. Then $N_{2\pi} = I$, from which it follows that the proper values of the diagonalizable generator n of this one-parameter group are integral. The other properties which occupation numbers notably should have are: the total momentum of the field (for any given type of momentum, forming a generator of the fundamental symmetry group, as usual) should be the sum of the products of the occupation numbers with the indicated momenta; the occupation numbers should transform in the obvious manner under a kinematical unitary transformation. In addition, the occupation numbers must be non-negative, or an interpretation in terms of anti-particles must be available; and in the familiar formalism, the occupation numbers are self-adjoint. While the first pair of these desiderata is satisfied, the second is not necessarily satisfied. While it is not entirely clear how serious the latter requirements are from a physical standpoint, it is of interest to inquire what properties a state of a Weyl algebra must have so that when interpreted as a physical vacuum, all of the particle-interpretation features mentioned are valid.

The self-adjointness of the occupation numbers is equivalent to the invariance of the physical vacuum under the automorphisms of the Weyl algebra induced by the cited phase transformations in the underlying (i.e., so-called "single-particle") unitary space. This is a strong condition, but is insufficient to fix the vacuum uniquely as the conventional free field vacuum. In fact there exist continuum many regular states which are Lorentz invariant and for which the corresponding occupation numbers are self-adjoint; these states are even invariant under all the automorphisms of the Weyl algebra induced by the unitary operators on $\mathscr{H}$.

The non-negativity requirement, however, is much stronger than might appear; it suffices to pick out the conventional free vacuum. A somewhat similar result applies to more general algebras of observables associated with a Hilbert space $\mathscr{H}$ (e.g., to Fermi-Dirac as well as to Bose-Einstein systems). To state this, suppose there is given for every unitary operator U on $\mathscr{H}$ (in physical terms, for every single-particle motion) a corresponding unitary operator $\Gamma(U)$ on a Hilbert space $\mathscr{K}$ (i.e. a corresponding "field" motion), such that Γ forms a continuous representation of the unitary group on $\mathscr{H}$. Assume also that the infinitesimal generator $d\Gamma(T)$ of the one-parameter unitary group $[\Gamma(e^{itT}); -\infty < t < \infty]$ is non-negative when T is non-negative. (The non-negativity of the occupation numbers is the special case of this when T is a projection; but in a formal way this special case is equivalent to the general case, in view of the formal expression available for any non-negative self-adjoint operator as a non-negative linear combination of projections.) Then $\mathscr{K}$ is a direct sum of irreducible representation spaces of the essentially classical type, i.e. the spaces of tensors of various symmetry types over $\mathscr{H}$, as in the work of Schur for finite-dimensional spaces $\mathscr{H}$, suitably topologized and completed in the case if infinite-dimensional $\mathscr{H}$. The Weyl algebra case is just that in which exclusively symmetric tensors are involved.

The main effect of these results is to confirm the validity of the first approach to the physical vacuum given in this chapter. In general, invariance requirements do not suffice to fix it, but non-negativity requirements on either the energy or the occupation numbers, together with rather elementary invariance requirements, will at least suffice to pick out the free vacuum, and in the case of the energy may reasonably be expected to apply in rather more complicated cases.

The free vacuum has played a part in the preceding discussion, as a special case in which to experiment with general notions about a physical vacuum, and is important in the mathematical and physical development, so that it is well to be explicit about it. It may be convenient as well as in analogy with common heuristic physical practice to speak of the algebra $\mathscr{A}$ as being "clothed" by the designation of a particular vacuum. The simplest vacuum is that which is completely unclothed, from a common physical outlook; in various contexts it is described as the "bare," "free," or "Fock-Cook" vacuums (the last name after the physicist and mathematician who first set it up in a clear-cut way). The rough idea is that the imposition of a non-trivial dynamics displaces the vacuum (i.e. it is no longer invariant under the temporal development), and so "dresses" the field, the new (physical) vacuum relative to the original (bare) vacuum consisting of a kind of "cloud" which "clothes" it. At any rate, the free vacuum and various mathematical structures associated with it play an important part in conventional theoretical physics and one can not expect to deal effectively with more general varieties of physical vacuums before this particularly accessible one has been explored. The next chapter will deal with the free vacuum, as well as some interesting related purely mathematical work (such as a variety of analysis in function space) and it may be useful to conclude this chapter with an indication of how it may relate to the vacuum of a specific interacting field, which will provide incidentally an example of non-trivial clothing.

Actually, the "free physical" field is regarded in theoretical physics as "clothed" relative to the "bare" field. Since empirical observation can be made only when there is interaction, the "bare" field is considered to be a purely mythical object by physicists, except as a mathematical entity. On the other hand, the observation of apparently free physical particles is an experimental commonplace and physicists commonly assume there is a free physical field closely related to reality. Since from a purely mathematical point of view this free physical field is taken to be identical with a bare field (with suitable physical parameters, e.g. the empirical physical mass will be used in the underlying field equations), the free field is mathematically as trivial as a bare field although it is "clothed" in the sense indicated.

To try to limit some of the confusion that tends to arise between "bare" and "clothed" fields, we shall use the term "zero-interaction" vacuum or field, generally in a purely mathematical way.

In the presence of interaction, the vacuum will differ from the zero-interaction vacuum and from one point of view this is the whole problem of field theory, virtually; but nevertheless the vacuum of the interacting field and that of the non-interacting one are in several ways closely related.

To set up a simple but relevant case, let $\mathscr{H}$ be a unitary space, $\mathscr{A}$ the Weyl algebra over $\mathscr{H}$, and E the universal vacuum described above. Let γ_0 be any inhomogeneous symplectic automorphism of $\mathscr{A}$. If ζ_t is the "free" motion of $\mathscr{A}$, or mathematically simply a one-parameter group of automorphisms of $\mathscr{A}$ inducible by a continuous one-parameter unitary group in the "free-field" representation, then $\zeta_t' = \gamma_0\zeta_t\gamma_0^{-1}$ is a new motion which can be rather thoroughly analyzed. In particular, its vacuum state is E^{γ_0}, i.e. the result of the contragredient action of γ_0 on E. This will in general be quite different from E; in fact, it may well happen that E^{γ_0} is not representable by any state vector in the free-field representation.

Any attempt to determine a state vector for E^{γ_0}, such as is made in conventional practice, will necessarily lead to a "divergent" expression, such as those involving infinite constants. That such an expression may be manipulated so as to arrive at finite, meaningful results, is not surprising, since it is in essence simply the rather ponderous result of translation of the simple state E^{γ_0} into an inappropriate and analytically untenable but algebraically tolerable framework.

A simple example of this is provided by a familiar soluble model considered by van Hove and others, which is supposedly representative of the key (so-called "ultraviolet") divergences of quantum field theory. To present this example in a quite elementary way, let us be initially heuristic and write the free and interaction hamiltonians simply as

$$H_0 = (1/2) \sum_k (c_k^2 p_k^2 + d_k^2 q_k^2) \quad (c_k > 0;\, d_k > 0;\, k = 1, 2, \cdots),$$
$$H_1 = \sum_k (a_k p_k + b_k q_k),$$

in terms of canonical variables $p_1, q_1, p_2, q_2, \cdots$. In the conventional field theory of several years ago, the "bare" vacuum is the lowest eigenstate of the "free-field" energy H_0, while the "physical" vacuum is the lowest eigenstate of the total hamiltonian $H_0 + H_1$ (i.e. the eigenvectors of least eigenvalue). The fundamental difficulty in quantum field theory, from one common outlook, is that although H_0 has a perfectly good interpretation as a self-adjoint operator, H_1 has remained apparently meaningless as an operator, without significant mathematical change since operators of a similar type were introduced in the first published work on quantized field theory (Dirac (1927)). People have tried in many ways to make sense out of H_1, but it is only with "renormalization," a partially ad hoc removal of infinite terms, that even quite partial results in this direction were attained.

What van Hove did was to show (for the particularly simple case when the a_k and b_k were scalars, and not operators as in physically more complicated cases) that conventional field-theoretic methods led to still another indication of the lack of mathematical meaning of H_1: the physical vacuum state representative was orthogonal to all eigenstates of H_0. There have been a number of paradoxes of this type (notable among them is one of Haag) and some people have even questioned the validity of the conventional foundation of quantum phenomenology in terms of Hilbert space.

To perform the indicated subsumption it is necessary, in view of the purely figurative existence, at this stage, of H_1, to work in a partially formal manner. What is mainly involved is the transformation of the total hamiltonian by completion of the square,

$$(1/2)(c_k^2p_k^2 + d_k^2q_k^2) + (a_kp_k + b_kq_k) = (1/2)c_kd_k(p_k'^2 + q_k'^2) + \text{const.},$$

where

$$p_k' = (c_k/d_k)^{(1/2)}p_k + (a_k/c_k)(c_kd_k)^{-(1/2)},$$
$$q_k' = (d_k/c_k)^{(1/2)}q_k + (b_k/d_k)(c_kd_k)^{-(1/2)}.$$

The p_k' and q_k' $(k = 1, 2, \cdots)$ form a new canonical system, related to the original one by an inhomogeneous symplectic automorphism; to see this it suffices to examine the one-dimensional case, in view of the lack of cross-terms between different indices, i.e. such a transformation as

$$p \to ap + r, \qquad q \to a^{-1}p + s.$$

The corresponding z's (elements of the phase space) may be taken to have the form $z = (u, v)$, u and v being real numbers here, with $R(z)$ taken as (the closure of) $up + vq$, and the fundamental form: $B[(u, v), (u', v')] = uv' - u'v$. The inhomogeneous symplectic automorphism

$$R((u, v)) \to R((au, a^{-1}v) + (ru + sv)$$

then carries p and q into the indicated new canonical variables.

The relevant formal development is now complete, and we may proceed to the rigorous formulation of the free and physical fields in question here. Let $\mathscr{M}$ be the space of all ordered pairs z of infinite sequences of real numbers, each sequence having the property that all the numbers vanish from some point onward:

$$z = (u_1, u_2, \cdots; v_1, v_2, \cdots) \quad (u_k = v_k = 0 \text{ for sufficiently large } k).$$

Define $\langle z, z'\rangle = \sum(u_kv_k' + v_ku_k') + i\sum_k (u_kv_k' - u_k'v_k)$. Let $\mathscr{A}$ be the corresponding algebra of field observables, and let $\mathscr{A}_0$ denote the algebra of all bounded functions of finite sets of the p's and q's. We first define the free vacuum E on the subalgebra $\mathscr{A}_0$. Let X be an arbitrary element of $\mathscr{A}_0$, say a function of $p_1, \cdots, p_n$ and $q_1, \cdots, q_n$. Let $p_1', \cdots, p_n', q_1', \cdots, q_n'$ denote, respectively, the operators

$$-i\partial/\partial x_1, \cdots, -i\partial/\partial x_n, \qquad x_1, \cdots, x_n,$$

formulated as self-adjoint operators in the usual way in the Hilbert space $L_2(E_n)$ of all square-integrable functions on n-dimensional euclidean space. There then exists, by the corollary to the Stone-von Neumann theorem described above, a unique algebraic isomorphism of the ring of all bounded functions of the $p_1, \cdots, p_n, q_1, \cdots, q_n$ with the corresponding ring for the primed set of canonical variables. This correspondence will take X into a well-defined operator X' on $L_2(E_n)$. We now set

$$E_0(X) = (X'v, v),$$

where v denotes the function on E_n,

$$v(x) = (\pi)^{-n/4} \exp\left[-\sum_{k=1}^{n} \frac{d_k}{c_k} x_k^2/2\right] \times \prod_{k=1}^{n} \left(\frac{c_k}{d_k}\right)^{-1/4}$$

Now it is a familiar fact that v is the lowest eigenfunction of $c_1^2p^2 + d_1^2q^2$ in one dimension, p and q being the usual operators in $L_2(-\infty, \infty)$. It follows in particular that E_0 is invariant under the one-parameter group generated by H_0, whose action is determined by its action on the elements of $\mathscr{A}_0$, this being:

$$X \to \exp\left[it \sum_1^n (c_k^2p_k^2 + d_k^2q_k^2)/2\right] X \exp\left[-it \sum_1^n (c_k^2p_k^2 + d_k^2q_k^2)/2\right].$$

In addition, it is readily verified that E_0 is uniquely defined, linear, positive, and normalized on $\mathscr{A}_0$, and it is immediate that $|E_0(X)| \leqq \|X\|$, $\|X\|$ denoting the bound of the operator X. It follows that E_0 extends uniquely to a state E of $\mathscr{A}$ that is invariant under H_0.

We have thus set up in a rigorous way the bare vacuum. To deal similarly with the physical vacuum, let θ denote the inhomogeneous symplectic automorphism of A which carries

$$R(z) \to R(Tz) + f(z),$$

where

$$T: z = (u_1, u_2, \cdots; v_1, v_2, \cdots) \to (r_1u_1, r_2u_2, \cdots; r_1^{-1}v_1, r_2^{-1}v_2, \cdots),$$

$$f(z) = \sum_k (u_ks_k + v_kt_k),$$

$$r_k = (c_k/d_k)^{(1/2)}, \quad s_k = (a_k/c_k)(c_kd_k)^{-(1/2)}, \quad t_k = (b_k/d_k)(c_kd_k)^{-(1/2)}.$$

Then the physical vacuum E' is *defined* by the equation

$$E'(X) = E^{\gamma_0^{-1}}(X),$$

i.e. as the state into which E is carried by the induced (contragredient) action on the states of canonical transformation (= automorphism of the algebra of field observables) γ_0. This physical vacuum is a mathematically rigorous object, devoid of divergences, and arrived at in a simple if slightly sophisticated manner.

How is it then that from the point of view of older conventional field theory this model is divergent? The answer is that it was tacitly effectively assumed that the physical vacuum should arise from a state vector, in the same representation as that in which the free-field vacuum was such. In fact, no calculus of states was employed except that based on the representation of states through vectors. Thus the bare vacuum was taken to have the form

$$E(X) = (X\psi_0, \psi_0),$$

while the physical vacuum was to have the form

$$E'(X) = (X\psi_0', \psi_0'),$$

ψ_0 and ψ_0' both being "unit" vectors in the underlying state vector space, say $\mathscr{K}$. Van Hove's paradox is then that ψ_0' can be formally demonstrated to be orthogonal to the eigenvectors of H_0, which on the other hand span $\mathscr{K}$.

The mathematical fact is simply that there exists no such vector ψ_0' in $\mathscr{K}$. This is connected with the non-implementability of the canonical transformation θ by a unitary operator in the zero-interaction representation. If such a unitary operator existed, say Γ (figuratively), then one could take simply $\psi_0' = \Gamma\psi_0$, and in terms of concrete operators on K, one would have

$$E'(X) = (X\Gamma\psi_0, \Gamma\psi_0) = E(\Gamma^{-1}X\Gamma).$$

(It may be helpful to note that formally one would have

$$\Gamma^{-1}(H_0 + H_1)\Gamma = H_0;$$

transformation by Γ of the fact that ψ_0 is the lowest eigenvector of H_0 then shows that $\Gamma\psi_0 = \psi_0'$ is the lowest eigenvector of $H_0 + H_1$.) However, it is mathematically demonstrable (see the cited references) that Γ does not exist. Nevertheless the purely figurative expression $\Gamma^{-1}X\Gamma$ can be given a rigorous mathematical interpretation, and E' can be well-defined, as indicated above. The mathematical non-existence of Γ suggests, but does not in itself establish, the non-existence of a vector ψ_0' in K such that E' will have the form

$$E'(X) = (X\psi_0', \psi_0');$$

this however is mathematically demonstrable. In this connection a quite general result of Shale is noteworthy. It asserts, roughly, that if T is any homogeneous symplectic transformation on a complex Hilbert space, which is not of the form UV, with U unitary and $V = I + W$, with W essentially Hilbert-Schmidt in a certain sense, then the transform of the associated bare vacuum by the induced action of T is non-normalizable in the zero-interaction representation.

We should perhaps explain that a state E is said to be normalizable if it arises from a state vector ψ in the fashion familiar in elementary quantum mechanics, $E(X) = (X\psi, \psi)$. This is of course a property not only of the state and algebra of observables, but also of the representation of the observables by concrete operators on a Hilbert space. As noted earlier, every state is normalizable in *some* representation. Non-normalizable states are familiar in elementary quantum mechanics in connection with the continuous spectrum. If T is a self-adjoint operator in a Hilbert space, and λ is a point of the continuous spectrum of T, then there exists (quite rigorously) a state of the system of all bounded operators, in which T has the exact value λ; but this state can not be normalizable. In a heuristic fashion this state may be represented in terms of a non-normalizable eigenvector for the eigenvalue λ, as is rather familiar. It is also familiar that such non-normalizable states are sometimes convenient to use for intuitive physical purposes, and can be dealt with in a mathematically and physically conservative way through the use of eigenvector packets. In the present case the non-normalizability has no such connection with the continuous spectrum of any operator (the total energy, of which ψ_0' is in the older conventional figurative sense an eigenvector, has discrete spectrum) and packets cannot be used effectively. However, the von Neumann formulation of states as linear functionals is applicable in even more effective form than in the case of a system of a finite number of degrees of freedom. The present physical vacuum state E', although non-normalizable in the zero-interaction representation, is nevertheless "regular," i.e. mathematically quite smooth; in particular, $E'(e^{iR(z)})$ is a continuous function of z relative to any finite-dimensional subspace of variation for the point z of the classical phase space. For systems of a finite number of degrees of freedom, the states which arise from the continuous spectra of observables are in contrast not regular and in particular $E'(e^{iR(z)})$ need not be an everywhere continuous function of the point z in phase space.

By virtue of the mutual correspondence between states and representations of C^*-algebras already mentioned, the physical vacuum state E' will determine a Hilbert space $\mathscr{K}'$, a concrete set of self-adjoint canonical variables over the classical phase space $\mathscr{M}$ which are represented in $\mathscr{H}'$, and other elements of structure, including notably a vector ψ_0' in $\mathscr{H}'$ such that

$$E'(X) = (X'\psi_0', \psi_0');$$

here X' denotes the concrete operator on $\mathscr{H}'$ corresponding to the element X of the algebra of field observables. This vector ψ_0' is perfectly normalizable, etc., but it is not in the original Hilbert space $\mathscr{H}$. The spaces $\mathscr{H}$ and $\mathscr{H}'$ are quite distinct; an isomorphism preserving the relevant elements of structure does not exist, due precisely to the non-normalizability of E' in the zero-interaction representation.

The apparently mathematically meaningless, figurative, operator $H_0 + H_1$ can now be interpreted as a self-adjoint operator in $\mathscr{H}'$, in the strict mathematical sense. This operator arises as the generator of the one-parameter group of unitary transformations in $\mathscr{H}'$ which are induced by the one-parameter group of automorphisms of the field observables obtained by transforming by γ_0 the free-field dynamics.

This has been a rather lengthy interpolation, but it should serve to show the importance of the zero-interaction vacuum state and representation even for the treatment of interacting fields, as well as to indicate that some interacting fields which are indubitably divergent from the older conventional standpoint can now be formulated in a mathematically unexceptionable way, which is at the same time somewhat more directly physical. It is appropriate now to enter into a rigorous and invariant account of the mathematics of the free vacuum.

References to Chapter V

See Segal (1959b and 1961), and the literature cited there for the main sources of the present chapter. The van Hove model cited is treated in his 1952 paper. His later more sophisticated models, developed from this original one and the Lee model, derive expressions for the physical vacuum and single-particle states which may well be rigorously explicable by a variant of the method of the present chapter.

CHAPTER VI

Representations of the Free Field

Various non-commuting operators occur in the theory of free fields and, depending on which ones are of special relevance, any one of three different representations, each of which diagonalizes certain of these operators, or otherwise permits their representation in simple form, may be advantageous. These representations may be described as: (1) the renormalized Schrödinger representation, an infinite-dimensional adaptation of the conventional Schrödinger representation (in this the field operators have a simple form, the kinematical transformations a fairly simple form, and the occupation numbers a relatively obscure form); (2) the Fock-Cook representation (in this the occupation numbers, as well as the kinematical transformations, are simple, while the field operators are relatively complicated); (3) the holomorphic functional representation (in this the creation and annihilation operators, as well as the kinematics, are simple, while the occupation numbers and the fields are less so, and the physical interpretation of the wave functions relatively difficult). All of these representations are naturally unitarily equivalent, essentially in canonical ways. The only one which appears in fairly clear-cut, if mathematically heuristic, form in the conventional literature, is the Fock-Cook representation. However, what we have called the renormalized Schrödinger representation can be regarded as a covariant description of an essentially familiar, if mathematically vague, description of the states of the free field in terms of infinite products of hermite functions. In addition, the quasi-canonical equivalence between these two representations can be regarded as a more explicit and covariant formulation of what Dirac describes as the representation of an assembly of bosons as a set of harmonic oscillators.

The adaption of the Schrödinger representation to the infinite-dimensional case depends on the development of analysis in function space along lines initiated by Wiener. The Wiener space itself, however, is not especially convenient for or relevant to field-theoretical applications, especially where covariance questions are involved. It is convenient rather to make the analysis directly in the Hilbert space of classical states.

Let $\mathscr{H}$ be a real Hilbert space. If $\mathscr{H}$ were finite-dimensional, the Schrödinger representation, with the suppression of the factor $\hbar$ in the definition of the canonical P's, yields a finite set of operators satisfying the canonical commutation relations. The infinite-dimensional case resisted a parallel treatment for some time because of the apparent lack of a suitable analogue to ordinary Lebesgue measure for Hilbert space. In fact, people proved that nothing approaching a halfway decent measure, in the conventional Lebesgue sense, could exist on a Hilbert space; to a literal-minded adherent to the classical Lebesgue formulation, such work as that of Feynman resembled a variety of scientific poetry devoid of apparent mathematical substance.

It turned out actually that when the real needs for and uses of integration and harmonic analysis in Hilbert space were kept firmly in mind, an effective and simple development of analysis over Hilbert space (i.e. analysis analogous to that of functions on euclidean space, except that the finite-dimensional euclidean space is replaced by a Hilbert space) was quite possible. The novelty of the integration could be assuaged by a reduction of it, if desired, to integration of the conventional variety over Wiener space or any of various related spaces of stochastic processes. But with the use of newer ideas about integration of a more operational character, it was a simple matter to give a direct formulation of the relevant integration theory, which is useful also in treating general representations of the Weyl relations. As far as applications to field theory go, there appears to be about as much need to introduce a special space on which to hang a countably-additive Lebesgue-like version of the virtual Hilbert space measure as to introduce an ether in connection with the Maxwell equations. On the other hand, the Wiener space is directly relevant to questions concerning Brownian motion, whose connection with the Laplacian makes analysis in the space useful in connection with the non-relativistic treatment of systems of a finite number of degrees of freedom whose hamiltonians involve the Laplacian.

To arrive at a notion of integration in a linear space of possibly infinite dimension, one may generalize the notion of probability distribution on the space. A probability distribution on a linear space $\mathscr{L}$ makes any linear function over $\mathscr{L}$ into a measurable function, with respect to the given distribution, and thus determines a linear map F from the dual $\mathscr{L}^*$ to random variables (=measurable functions, essentially) on a measure space (some elementary topological restrictions being required in the case when $\mathscr{L}$ is infinite-dimensional). Conversely, any such map determines a probability distribution on $\mathscr{L}$ uniquely, when $\mathscr{L}$ is finite-dimensional. For $\mathscr{L}$ infinite-dimensional there are however such linear maps, indeed such of quite a regular sort, that do not arise from any conventional probability distribution on $\mathscr{L}$. In fact, in quantum field theory as well as in the theory of stochastic processes the relevant distributions are frequently of this type.

For example, if $x(t)$ denotes the conventional Brownian motion process on the time interval $0 \leqq t \leqq 1$, and $\mathscr{L}$ is $L_2(0, 1)$, then the map

$$f(t) \to \int_0^1 f(t)\, dx(t)$$

where $f(t)$ denotes the function which determines an element of the dual $\mathscr{L}^*$ of $L_2(0, 1)$ (i.e. the element taking a general g in $L_2(0, 1)$ into $\int_0^1 g(t)f(t)\, dt$), is a distribution in the indicated generalized sense, which does not arise from any countably additive measure on subsets of $\mathscr{L}$. The integral in question is the Wiener stochastic integral and the mapping in question is precisely that which establishes the connection between integration in a real Hilbert space and in Wiener space.

Now defining a distribution on a topological linear space $\mathscr{L}$ as a linear map from the dual $\mathscr{L}^*$ of $\mathscr{L}$ to random variables on any probability space (the conventional type of distribution given by a completely additive measure is relatively less important in the case of an infinite-dimensional space and may be referred to more logically by a special term which we do not need to introduce as we shall have no occasion to refer to it), this definition is of course effective only if one can then integrate suitable functions on $\mathscr{L}$. Now if $\mathscr{A}$ is the collection of all functions on $\mathscr{L}$ that depend continuously on a finite number of linear functionals, and are bounded, then it is evident that $\mathscr{A}$ is an algebra and it is clear how the expectation of any element f of $\mathscr{A}$, say $E(f)$, should be defined, if one thinks about it for a moment. This structure $(\mathscr{A}, E)$ consisting of an algebra with a distinguished linear functional on it is easily seen to have certain properties which are known to characterize (weakly) dense subalgebras of the algebra of all bounded measurable functions on a (countably additive) measure space, together with the integral as functional E, this measure space being essentially unique (apart from special labelling of the points or sets). The Lebesgue theory is thereby applicable. Actually, the Lebesgue theory may be readily dispensed with and the whole theory built up on the basis of an algebra-cum-positive linear functional. In either way the space $L_2(\mathscr{L})$ acquires definite meaning and becomes a viable Hilbert space although not all elements of it are represented by actual functionals on $\mathscr{L}$—this is the main difference between the present situation and that involving integration with respect to a countably additive measure on a measure space.

Any $\mathscr{L}$ is the inverse limit of quotient spaces in such a way that the integral over L is a limit of countably additive integrals on these spaces. Alternatively, the integral can be derived from a finitely additive measure on $\mathscr{L}$ which is countably additive on the subsets invariant under translation by the vectors in any fixed kernel subspace, —or from countably additive measures on each finite-dimensional subspace, in the case of a Hilbert space. This approach is superficially closer to a conventional measure-theoretic one, but is not particularly helpful otherwise.

To give an example, let $\mathscr{H}$ be a real Hilbert space. Among the simplest interesting properties a distribution n might have which connect with the Hilbert space structure are: (1) unitary invariance, i.e. the joint distribution of $n(x_1), \cdots, n(x_k)$ and of $n(Ux_1), \cdots, n(Ux_k)$, for any unitary operator U and vectors $x_1, \cdots, x_k$ in $\mathscr{H}$ (which is henceforth canonically identified with its dual—so we may take a distribution on a Hilbert space as being a linear map on the space itself) are identical, and (2) equivalence of orthogonality and stochastic independence, in the sense that two mutually orthogonal sets of vectors in $\mathscr{H}$ are taken by n into two stochastically independent sets of random variables. It turns out that such a distribution exists and is essentially unique (a fact due essentially to Kac for the basic case in which $\mathscr{H}$ is two-dimensional). It may be called the isotropic centered normal distribution: for any vector x in $\mathscr{H}$, $n(x)$ is normally distributed with mean 0 and variance $c|x|^2$.

In the case of a finite-dimensional space, analysis based on ordinary euclidean measure may be readily seen to be equivalent to analysis based on the centered isotropic normal distribution; the point is that the two distributions are mutually absolutely continuous. In the infinite-dimensional case there is no direct analogue to ordinary euclidean measure, but one can get good counterparts to most of the global theorems in euclidean analysis in terms of the normal distribution. Actually it is simpler for some purposes to work in terms of the normal distribution even in the finite-dimensional case.

As an illustration of how finite-dimensional results may be adapted to the infinite-dimensional case, consider the Plancherel theory. Formally physicists frequently write such integrals as $\int_H e^{i(x,y)}f(x)dx$

over a Hilbert space $\mathscr{H}$. Such an integral has merely figurative existence, but one can make a certain "infinite renormalization" and replace the Fourier transform by a mathematical rigorous Wiener transform, which is rigorously equivalent to the Fourier transform in the finite-dimensional case and formally a substitute for it in the infinite-dimensional case. Specifically, one has the following theorem:

The transformation on the algebra $\mathscr{P}$ of all polynomials over $\mathscr{H}$,

$$f(x) \to \int_{\mathscr{H}} f(2^{1/2}x + iy)\, dn(y)$$

extends uniquely to a unitary transformation on $L_2(\mathscr{H}, n)$ whose inverse on $\mathscr{P}$ is

$$F(x) \to \int_{\mathscr{H}} F(2^{1/2}x - iy)\, dn(y).$$

In the finite-dimensional case one readily deduces from this the Plancherel theorem, essentially by completing the square in the exponentials involved. The Wiener transform defined above has the distinctive special property of mapping polynomials into polynomials. It should perhaps be mentioned that a polynomial on a Hilbert space is simply a polynomial in the usual sense in a finite number of coordinates on the space. A similar transform was established by Cameron and Martin for functionals on Wiener space at a relatively early date. Through the connection of Wiener space with a real Hilbert space the relationship to the conventional Fourier transform is visible.

A good deal more might be said about analysis over a Hilbert space, but for brevity and because some of this is implicit in the formulation of the renormalized Schrödinger representation, we simply state the form of the canonical variables in this representation.

For any vector x in $\mathscr{H}$, let $Q(x)$ and $P(x)$ denote the self-adjoint generators of the one-parameter unitary groups $\{A(tx);\ -\infty < t < \infty\}$ and $\{B(tx);\ -\infty < t < \infty\}$ respectively, where these operators act on $L_2(\mathscr{H})$ as follows:

$$A(u): f(x) \to \exp[i(x, u)/2^{1/2}] f(x),$$
$$B(u): f(x) \to \exp[-(x, u)/2^{1/2} - (u, u)/2^{1/2}] f(2^{1/2}u + x).$$

Then the canonical commutation relations hold, in the Weyl form:

$$\exp[-iP(x)] \exp[iQ(y)] \exp[iP(x)] \exp[-iQ(y)] = \exp[-i(x, y)]I.$$

The distribution on H is the canonical normal one with variance parameter $c = 1$ (other values of c give materially different canonical variables). In the finite-dimensional case the foregoing representation for the canonical variables is unitarily equivalent to the Schrödinger one via the trivial operation of multiplication by a constant multiple of $\exp[-(u, u)/4]$. Strictly speaking the $A(u)$ and $B(u)$ above are first defined on bounded continuous functions of a finite number of linear functionals (coordinates), and then extended by continuity to all of $L_2(H)$. The parallelism with the theory in the finite-dimensional case may be illustrated by the fact that, just as the Fourier transform then takes $P(x)$ into $Q(x)$ and $Q(x)$ into $-P(x)$, with the Schrödinger P's and Q's, the Wiener transform does the same for the present P's and Q's, irrespective of dimension.

To lead up to the connection between the representation just formulated and the Fock-Cook representation, observe that because of the orthogonal invariance of the canonical normal distribution, there is for any orthogonal transformation V on $\mathscr{H}$ a corresponding unitary transformation $\Gamma(V)$ on $L_2(\mathscr{H})$: $f(x) \rightarrow f(V^{-1}x)$ (here again the indicated operation must first be defined on the "nice" functionals described previously, and then extended to all of $L_2(H)$). Now it is noteworthy that this representation may be extended, without enlargement of the representation space, to the unitary operators on the "complex extension" of $\mathscr{H}$, i.e. the complex Hilbert space $\mathscr{H}'$ indicated by the notation $\mathscr{H}' = \mathscr{H} + i\mathscr{H}$. The simplest such unitary operation, that of multiplication by i, turns out to be, e.g., just the Wiener transform. Now the reduction into irreducible constituents of this extended representation is from a physical viewpoint just the reduction of the state vector space of the field into the direct sum of the n-particle subspaces; from a probabilistic point of view not involving fields and particles, virtually the same reduction occurs, in the Wiener space representation, in Wiener's well-known homogeneous chaos paper. It turns out that for each non-negative integer k, there is a subspace $\mathscr{M}_k$ of $L_2(\mathscr{H})$, which transforms under the action of the $\Gamma(U)$, just like the (covariant) symmetric tensors over $\mathscr{H}'$ of rank k; and this representation, in the space $\mathscr{H}'_k$ of symmetric k-tensors, is irreducible, as it is well-known to be in the finite-dimensional case. The infinitesimal generator of the one-parameter unitary group $\{\Gamma(e^{it}); -\infty < t < \infty\}$ is the operator known physically as the "total number of particles operator," and the most general bounded operator commuting with all $\Gamma(U)$, U unitary, is simply a function of this particle operator N. The action of N on its eigenmanifold equivalent to $\mathscr{H}'_k$ is, as the terminology indicates, merely multiplication by k.

This leads to an explicit equivalence between the renormalized Schrödinger, or "wave" representation, and the Fock-Cook, or "particle" representation, which corresponds to the wave-particle duality in a quantum field. To describe the latter representation explicitly, we define a *covariant tensor of rank k*, over a complex Hilbert space $\mathscr{H}$, as a multi-linear functional ϕ on the k-fold direct sum of copies of the dual of $\mathscr{H}^*$ with itself

$$(x_1^*, \cdots, x_k^*) \rightarrow \phi(x_1^*, \cdots, x_k^*);$$

a *symmetric* tensor as one satisfying the condition

$$\phi(x_{(1)}^*, \cdots, x_{(k)}^*) = \phi(x_1^*, \cdots, x_k^*)$$

for any permutation of the indices $1, \cdots, k$; and *square-integrable* (or *normalizable*) tensor as one satisfying the boundedness condition

$$\sum_{i_1, i_2, \cdots, i_k} |\phi(e_{i_1}, e_{i_2}, \cdots, e_{i_k})|^2 < \infty,$$

when the e_i constitute a maximal orthonormal set in $\mathscr{H}$. It is demonstrable that if this boundedness condition holds for any one maximal orthonormal set, then it holds for any other, and in fact the inner product

$$(\phi, \phi') = \sum_{i_1, i_2, \cdots, i_k} \phi(e_{i_1}, \cdots, e_{i_k})\bar{\phi}'(e_{i_1}, \cdots, e_{i_k})$$

is independent of the choice of basis for square-integrable tensors. It then follows that the covariant square-integrable k-tensors over $\mathscr{H}$ form, in a canonical way, a Hilbert space relative to the stated inner product, of which the symmetric tensors form a closed subspace $\mathscr{H}_k$, the space of all symmetric covariant k-tensors over $\mathscr{H}$.

If for example $k = 1$, then this space, $\mathscr{H}_1$, is canonically identical with $\mathscr{H}$. For $k = 0$, it is convenient to define a tensor simply as a constant (complex number), and consider all 0-tensors as symmetric and square-integrable, and to formulate the corresponding space $\mathscr{H}_0$ as a one-dimensional Hilbert space relative to the inner product: $(a, b) = a\bar{b}$. The direct sum $\sum_{k=0}^{\infty} \mathscr{H}_k$ of all such tensors is then a Hilbert space $\mathscr{K}$ canonically associated with $\mathscr{H}$. For any unitary operator U on $\mathscr{H}$ it follows that there is a corresponding unitary $\Gamma(U)$ on $\mathscr{K}$; specifically, $\Gamma(U)$ takes the vector ϕ in $\mathscr{H}_k$ into ϕ', where

$$\phi'(x_1^*, \cdots, x_k^*) = \phi((Ux_1)^*, \cdots, (Ux_k)^*),$$

x^* denoting for any vector x in $\mathscr{H}$ the linear functional: $y \rightarrow (y, x)$.

There is no difficulty in verifying that $\Gamma(UU') = \Gamma(U)\Gamma(U')$ for arbitrary unitary operators U and U' on $\mathscr{H}$, and that the map $U \to \Gamma(U)$ is continuous, so that Γ affords a continuous unitary representation on $\mathscr{K}$ of the full unitary group on $\mathscr{H}$. It might be mentioned parenthetically that Γ is a remarkable representation from a purely group-theoretic point of view, having in particular the property that if R_i is a representation of a group G on a Hilbert space $\mathscr{H}_i$ $(i = 1, 2)$, then $\mathscr{K}(\mathscr{H}_1 + \mathscr{H}_2)$ is canonically isomorphic with $\mathscr{K}(\mathscr{H}_1) \times \mathscr{K}(\mathscr{H}_2)$, so that $\Gamma(R_1(g) + R_2(g)) \cong \Gamma(R_1(g)) \times \Gamma(R_2(g))$. This means that, roughly speaking, Γ behaves like the exponential function for representations; and it might be of purely mathematical interest to determine all such universal representations satisfying the same functional equation, etc.

If x is any vector in $\mathscr{H}$, and x^* is the corresponding element of $\mathscr{H}^*$, the operation $\phi \to \phi'$ taking a k-tensor ϕ into the $(k + 1)$-tensor ϕ' given by the equation

$$(k + 1)!\phi'(u_1^*, \cdots, u_{k+1}^*) =$$
$$(k + 1)^{(1/2)} \sum_{\pi} (x, u_{\pi(l)})\phi(u_{\pi(1)}^*, \cdots, \hat{u}_{\pi(l)}^*, \cdots, u_{\pi(k)}^*),$$

(π ranging over the group of all permutations of $1, 2, \cdots, k + 1$), where the $\hat{}$ over the u_l^* signifies that this variable is deleted, has a unique linear extension to the algebraic direct sum $\mathscr{S}$ of the $\mathscr{K}_k$, which as an operator on the Hilbert space $\mathscr{K}$ admits a closure $C(x)$ called the "creation operator for an x-particle." Its adjoint $C(x)^*$, whose domain also includes $\mathscr{S}$, is called the "destruction operator for an x-particle." The nomenclature here derives from the fact that $C(x)$ increases the "number of x-particles," as defined above with the use of the representation Γ just defined, by one, while $C(x)^*$ decreases it by one when acting on a vector representing a state including at least one x-particle.

It is not difficult to check that the operators $C_0(x)$ and $C_0^*(x)$ obtained by restricting $C(x)$ and $C(x)^*$ to $\mathscr{D}$ satisfy the relations

$$[C_0(x), C_0(y)^*] = -(x, y),$$

from which it follows that if $R_0(x)$ is defined by the equation

$$R_0(x) = (C_0(x) + C_0^*(x))/2^{1/2},$$

then

$$[R_0(x), R_0(y)] = -i\,\mathrm{Im}(x, y).$$

This indicates that $R_0(\cdot)$ may be extendable to a set of generators for a Weyl system, and a rigorous analysis shows that this is in fact the case.

Actually, about the most economical way to establish the Weyl relations here is to set up the isomorphism of the tensor system just described with the generalized Schrödinger system described earlier and use the validity of the relations in that system. To describe this isomorphism, let $\mathscr{H}'$ be any real Hilbert space such that $\mathscr{H} = \mathscr{H}' + i\mathscr{H}'$, i.e. $\mathscr{H}$ is the complex extension of $\mathscr{H}'$. Any element of the dense subspace $\mathscr{S}$ of the normalizable symmetric tensors over $\mathscr{H}$ gives a simple functional on $\mathscr{H}^*$ when all the variables are set equal to a single variable x^*, in fact the resulting functional is a polynomial over $\mathscr{H}^*$. Its restriction to $\mathscr{H}'$ is therefore in the domain of the Wiener transform for an arbitrary variance parameter, in particular the parameter 1/2. Now transforming a given such k-tensor successively into the indicated functional on $\mathscr{H}^*$ and applying the Wiener transform on $\mathscr{H}'$ with variance parameter 1/2, and multiplying the result by $(k!)^{-(1/2)}$, the resulting polynomial on $\mathscr{H}'$ is defined as the corresponding functional in $L_2(\mathscr{H}')$ to the given symmetric tensor. (Note that the relevant distribution on $\mathscr{H}'$ has variance parameter 1, not 1/2.) It can be shown that this correspondence extends uniquely by linearity and continuity to a unitary transformation of the space $\mathscr{K}$ of all normalizable covariant symmetric tensors over $\mathscr{H}$ onto $L_2(\mathscr{H}')$, in such a manner that $R_0(z)$, as defined above, is carried into an operator in $L_2(\mathscr{H}')$ whose closure is the self-adjoint generator $R(z)$ of the Weyl system associated with that space in the manner indicated above.

There is another representation which is somewhat similar to the renormalized Schrödinger representation, but differs from it in that the representation space consists of functionals on the classical phase space $\mathscr{H}$, rather than a real part $\mathscr{H}'$ of $\mathscr{H}$, and attains irreducibility through restriction to holomorphic, rather than arbitrary square-integrable, functionals. This representation has not as yet been employed in a clear-cut way in the literature, but it clarifies the structure of the creation and annihilation operators, and may well be useful in later developments. The use of square-integrable functions over a phase space as a representation space for dynamical variables is due originally to Koopman, who investigated classical systems of a finite number of degrees of freedom by these means, but without introducing field variables or restricting to holomorphic functionals.

By a polynomial on a complex Hilbert space $\mathscr{H}$ is meant a function $p(\cdot)$ on $\mathscr{H}$ such that $p(z)$ is expressible as a polynomial in the elementary sense in a finite number of inner products $(z, e_1), \cdots, (z, e_n)$, the e_i being fixed vectors in $\mathscr{H}$. The representation mentioned can be effected canonically either on the anti-holomorphic functionals on $\mathscr{H}$

or the holomorphic functionals on the dual $\mathscr{H}^*$ of $\mathscr{H}$ (a holomorphic functional being a limit of polynomials in an appropriate sense). The correspondence between these two sets of functionals is the unique unitary one extending the assignment to any polynomial p on $\mathscr{H}^*$ the anti-polynomial (i.e. complex conjugate functional to a polynomial) p^* defined as follows: if $u \in \mathscr{H}$, and if u^* is the linear functional given by the equation $u^*(w) = (w, u)$, w being a generic element of $\mathscr{H}$, then $p^*(u) = p(u^*)$. It will suffice to describe the representation in the space of holomorphic functionals on $\mathscr{H}^*$.

Let n be the centered normal distribution of unit variance on $\mathscr{H}^*$ as a real space, and $\mathscr{K}$ the closure in $L_2(\mathscr{H}^*, n)$ of the set of all polynomials on $\mathscr{H}^*$. Let $W(z)$ for arbitrary z in $\mathscr{H}$ be the unitary operator on $\mathscr{K}$ uniquely determined by the property that it operates as follows on polynomials $p(u^*)$:

$$p(u^*) \to p(u^* + z^*) \exp[-(1/4)(z, z) - (1/2)u^*(z)],$$

z^* denoting as usual the linear functional determined by z, $u \to (u, z)$. It can then be shown from the transformation properties of the normal distribution under translations that $W(\cdot)$ is a Weyl system. For any unitary operator U on $\mathscr{H}$, let $\Gamma(U)$ now be defined as the unitary operator on $\mathscr{K}$ uniquely determined by the property that it acts as follows on polynomials:

$$p(u^*) \to p((Uu)^*).$$

There is no difficulty in verifying that Γ is a continuous unitary representation on $\mathscr{K}$ of the full unitary group on $\mathscr{H}$. If v^* denotes the functional identically one on $\mathscr{H}^*$, it can be shown that we have here the conventional free-field system, within unitary equivalence, with Γ as the representation transferring single-particle to field motions, and v^* as the vacuum state representative.

One of the notable features of this representation is the particularly transparent form in which the creation and annihilation operators appear. It is straightforward to compute the restriction of the field variables $R(z)$ which serve as generators for the unitary groups with which the Weyl relations are concerned and to derive from this the actions of $C(z)$ and $C(z)^*$ on the polynomials. The result is that, apart from factors of $\pm(-1/2)^{\pm(1/2)}$, the creation operator for a particle with wave function z acts as multiplication by (z, u), while the annihilation operator acts as differentiation in the direction z (i.e. the unique derivation on the algebra of polynomials on $\mathscr{H}^*$ taking $u^*(w)$, for any w in $\mathscr{H}$, into (w, z)).

The relation between the holomorphic functional representation just presented and the particle representation is a simple one. Any k-tensor f determines a functional on $\mathscr{H}^*$ by taking all the variables on which the value of f depends to be the same, and this transformation is, essentially, apart from constant factors, the unitary transformation mapping the one field representation into the other. How these various representations are used depends on the situation at hand. For example, to derive the spectrum of the flow associated with Brownian motion, the equivalence of the particle and the Schrödinger representation may be employed, yielding in a simpler though related way the results first obtained by Kakutani (1950). On the other hand, to make sense out of the notion of an exponential or other entire function of a creation or annihilation operator, a problem which arises if it is desired to represent exponentials or other functions of the hermitian field variables $R(z)$ in the so-called Wick normal form, the holomorphic functional representation is the only one that seems useful. It affords a type of pseudo-diagonalization for the non-diagonalizable (as well as unbounded) creation and annihilation operators and suggests an approach to such relatively intractable operators whose development may have purely mathematical interest.

References to Chapter VI

The Fock representation (1932) was the first explicit, if rather heuristic, formulation of a representation space for the operators of quantum field theory. Its rigorous mathematical explication by Cook (1953) provided the first rigorous formulation for quantum mechanical variables and proofs of such basic matters as their essential self-adjointness. The renormalized Schrödinger representation was treated by Segal (1956). In a general quantum-mechanical proposal, Siegel and Wiener (1953) stressed the use of unitary groups arising from groups of measure-preserving transformations, and noted how the work of Paley and Wiener yields a representation of the unitary group on $L_2(-\infty, \infty)$ by measure-preserving transformations on Wiener space. Rather remarkably, this representation is essentially equivalent to the representation Γ of the unitary group associated with Bose-Einstein quantization, with $L_2(-\infty, \infty)$ taken as the underlying Hilbert space, while in the case of Fermi-Dirac quantization, measure-preserving transformations on a "non-commutative measure space" arise (cf. Segal (1956b)). In the heuristic treatment of Friedrichs (1953) for Bose-Einstein quantization the representation Γ does not explicitly intervene, but an inner product for "Hermite functionals" on Hilbert space is defined in connection with the representation of field variables. As noted earlier, the duality between the Fock (particle) and renormalized Schrödinger (wave) representation has its roots in the familiar representation of a Bose-Einstein field in terms of an assembly of harmonic oscillators and their associated Hermite eigen-functions, which was treated at an early date by Dirac, Fermi, and others.

The holomorphic functional representation connects with a Weyl system first formulated by Shale, with representation space taken as all square-integrable functionals on a complex Hilbert space. This representation was not irreducible, in fact the transforms of the unit functional under the operators defining the Weyl system yield precisely the holomorphic functionals, as shown by Segal (1962). It is better adapted to the problem of non-linear relativistic quantization than the other representations, and plays an implicit role in Segal (1960b).

CHAPTER VII

Interacting Fields: Quantum Electrodynamics

The key mathematical question in connection with interacting fields is not so much one of the validity of various theorems in the usual mathematical sense, but rather of the effective formulation of the concept of an interacting field. To be effective, there must be non-trivial examples, but at the present time there are no non-trivial examples at all which are relativistic and involve particle creation, regardless of how the notion of interacting field is formulated. This does not prevent the "theory of interacting fields" from being successfully applied in non-trivial cases, such as relativistic quantum electrodynamics, but it means that it is not really a theory in precisely the usual mathematical physical sense. Rather, it is an assembly of relatively special methods and ideas, largely of a heuristic and even vague character, which have been able to deal with particular problems. This is not to say the subject is without mathematical beauty, for actually such could be regarded as the chief reason for its study; but real assurance that this beauty is not, in crucial aspects, illusory, is so far lacking.

Two mathematical approaches, the constructive and the axiomatic, seem possible, but there is not a sharp line between them since in either case a formulation, in part at least, of what a quantum field consists of, is required. A number of different formulations have been given, all of which can fairly be represented to codify aspects derived from existing thought and heuristic practice concerning quantum fields. Among these are e.g. the generating functional approach developed by Schwinger; the approach in terms of vacuum expectation values of products of fields studied intensively in recent years by Källén and Wightman; one along lines implicit in our earlier chapters, emphasizing the use of the so-called in- and out-fields, and utilizing the representation independent formalism; and there are many others, some of which are variants of the foregoing ones. It is quite impossible, in the nature of things, to prove in a strict mathematical sense that quantum electrodynamics or any other interacting field does or does not exist; the most that can be hoped for is that some reasonable set of analytic desiderata can be shown to be valid or invalid relative to various specific formulations.

This is a difficult context in which to seek definitive results and such seem only possible in a significant sense if rigorous and effective mathematical formulations which are physically conservative and fundamentally simple are available. The mathematical difficulties are apparent when one considers that, while there remain many serious difficulties in the global theory of non-linear hyperbolic equations quantum field theory is largely concerned with equations of that sort in which the dependent variable is not numerically valued, but is a pseudo-operator in an infinite-dimensional space—not an operator in the strict sense, not even an unbounded one, but a kind of germ of an operator.

The simple explication of the van Hove model of a divergent field given above suggests the application of similar methods to relativistic interacting fields. This can be done up to a point in a moderately conservative way. In this chapter we shall show, on the positive side, that in typical such cases, as e.g. quantum electrodynamics, the interaction hamiltonian becomes convergent if a suitable representation of the canonical variables is used. This refers to the formulation of the fields in the so-called interaction picture which seems in a way more closely related to measurement than the so-called Heisenberg picture, as well as probably closer to the ideas of renormalization in their simplest form. However, the interaction hamiltonian is then time dependent so that the question of whether there exists a single representation in which, for all times, the interaction hamiltonian is convergent, naturally arises. This is a difficult question but it can at least be reduced to a certain extent to well-defined mathematical questions which seem independent of variations in the details of the formulation. On the other hand, even if it is proved that such a representation exists or does not exist, i.e. a rigorous type of convergence or divergence is established, nothing definite can be asserted about the convergence or divergence of various formulations in terms of the Heisenberg picture. Roughly speaking, the difficulty seems to be that in field theory one is trying to describe a non-linear situation (the interacting field) in linear terms (utilizing either the "bare" field, or the asymptotic free physical field, both of which are linear); one must consequently expect a good deal of intricate complication in the results. How this difficulty might be remedied will be discussed in the next chapter.

The most basic interactions of field theory, the linear-bilinear interactions between Bose-Einstein and Fermi-Dirac fields, can be described in formal terms in various ways, by means of the equations of motion, in terms of the interaction hamiltonian, etc. None of these formulations has clear mathematical meaning, and for brevity we deal here with that in terms of the interaction hamiltonian. In the "interaction picture" this "operator" $H_I(t)$ is time dependent and has the form

$$H_I(t) \sim \int \boldsymbol{\phi}(x, t) \cdot \mathbf{j}(x, t)\, d_3x,$$

where $\mathbf{j}(x, t)$ is a hermitian-bilinear expression in the fermion field $\boldsymbol{\psi}(x, t)$; both $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ satisfy the free field equations here, this being precisely the formal advantage of the interaction representation. Now $\boldsymbol{\psi}(x, t)$ has no real mathematical meaning, but is the operator-valued quasi-function similar to that described in Chapter III, i.e. is the figurative kernel of the linear map $f \rightarrow \Phi(f) \sim \int \boldsymbol{\phi}(x) f(x) d_4x$, from suitable functions f on space-time to the operators on the Bose-Einstein state vector space. This is quite vague mathematically, but any more precise rendering would represent a special interpretation of the situation not incorporated in the generally accepted ideas on the subject. Of course, to make mathematical progress, a rigorous interpretation must be made, but this naturally involves the danger of becoming involved in difficult technical questions which may in the end turn out to be not quite relevant.

To arrive at a conservative mathematical interpretation, let us proceed to begin with in a formal way. To fix the ideas one may, if desired, think of $\boldsymbol{\phi}$ as a photon field $\boldsymbol{\phi}_\mu$ and $\boldsymbol{\psi}$ as an electron field $\boldsymbol{\psi}_\mu$, in which case $\mathbf{j}_\mu = \bar{\boldsymbol{\psi}} \cdot \gamma_\mu \boldsymbol{\psi}$, and $H_I(t) \sim \int (\sum_\mu \boldsymbol{\phi}_\mu \mathbf{j}_\mu) d_3x$. The general case proceeds along lines which are not, as regards present issues, significantly different. To say that $\boldsymbol{\phi}$ is a quantized photon field is formally equivalent to the possibility of making an expansion of $\boldsymbol{\phi}$ in terms of classical photon wave functions with operator coefficients satisfying the canonical commutation relations, say

$$\boldsymbol{\phi}(x) \sim \sum_{k=1}^{\infty} R_k \phi_k(x),$$

$\{\phi_k\}$ being an orthonormal basis for normalizable solutions of the Maxwell equations, and the R_k being the usual operators whose mutual commutators are scalars. Similarly, $\boldsymbol{\psi}$ may be expanded in classical solutions of the Dirac equation:

$$\boldsymbol{\psi}(x) \sim \sum_k S_k \psi_k(x),$$

the S_k having scalar anti-commutators. Now the given formal expression for H_I fails to make mathematical sense in the beginning because of the lack of mathematical meaning for such products as $\phi_\mu \bar{\psi}\gamma_\mu \psi$; this is not at all a quasi-function in the same sense as the ϕ or the ψ, and there is no clear-cut mathematical or physical way known to give such a product at a particular point a definite meaning. But let us substitute the expressions for $\phi(x)$ and $\psi(x)$ in terms of classical fields, still proceeding, necessarily, in a formal manner.

It results that

$$H_I(t) \sim \sum_{ijk} R_i S_j^* S_k \, c_{ijk},$$

where the constants c_{ijk} have the form

$$c_{ijk} = \int \phi_i(x) \bar{\psi}_j(x) \psi_k(x) d_3 x,$$

or

$$H_I(t) \sim \sum_i R_i N_i, \qquad N_i = \sum_{ijk} c_{ijk} \, S_j^* S_k.$$

This is an improvement over the original form for H_I, which was troublesome because it involved local products of fields, in that the individual operators R_i and N_i, and the finite sums $\sum_{i=1}^{n} R_i N_i$, can be given effective mathematical meaning, and the divergence of H_I resides only in the lack of an effective limit to the indicated sum, which is a more familiar kind of divergence. Let us therefore reformulate $H_I(t)$, as the indicated sum. To insure the finiteness of the coefficients c_{ijk} let us simply "put the system in a box," i.e. consider the space-time coordinates to be limited in a fixed but arbitrary way, with periodic boundary conditions imposed; since the box may be made arbitrarily large, any physical results should be approximable arbitrarily closely and mathematically the fundamental convergence questions can be expected to be unaffected by this procedure, which affords a relatively direct approach to these questions.

The box limitation also suffices to make the operators N_i perfectly good self-adjoint operators in the Fermi-Dirac free-field state vector space, although they will appear as divergent in some loose formulations. The point is that such bilinear expressions in fields as N_i are really formally equivalent to the mathematically unexceptionable object $d\Gamma(A_i)$, where Γ is the representation taking classical field motions into quantum field ones, $d\Gamma$ is the infinitesimal representation associated with Γ, and A_i is an honest self-adjoint operator in the classical underlying Hilbert space. The most familiar case of this is the expression $\int\psi^*(x)\psi(x)d_3x$ for the total number of particles in a common conventional heuristic approach; more properly it is expressed as $d\Gamma(I)$, I being the identity operator. It is not difficult to make explicit if somewhat tedious computations of the A_i in specific instances. The "dropping of vacuum expectation value terms" which are commonly "infinite constants" is not required because in the present formulation this vacuum expectation value is automatically zero.

Now there is one more formal aspect of the situation that is essential: the N_i are mutually commutative. This follows from a familiar feature of free-field quantization, the "micro-causality," to the effect that there is mutual commutativity at equal times, of the Bose-Einstein "field" values $\phi(x, t)$ (at different points x), and of the Fermi-Dirac current $j(x, t)$. These properties, while sometimes presented as postulates, are automatic consequences for free fields of the indicated covariant quantization rules. Since the N_i are linearly dependent on the $j(x, t)$ for fixed t, they also are mutually commutative, in a formal way. A rigorous argument can be given for each interaction by computation of the operators A_i such that $N_i = d\Gamma(A_i)$, and verification that the A_i, which are well-behaved operators on the underlying single-particle space $\mathscr{H}$ for the fermions, are mutually commutative. For since Γ is a representation, then for any commuting self-adjoint operators A and B on $\mathscr{H}$, it follows from the commutativity of the one parameter groups e^{itA} and $e^{itB}(-\infty < s, t < \infty)$ that the corresponding groups $\Gamma(e^{itA})$ and $\Gamma(e^{itB})$ commute, so that their self-adjoint generators $d\Gamma(A)$ and $d\Gamma(B)$ likewise commute in the strict Hilbert space sense.

At this point a rigorous definition of $H_I(t)$ may be made:

$$H_I(t) = \sum_i (R_i N_i)^{\sim},$$

where the R_i are canonical variables on the photon state vector space (in an as yet unspecified representation), and the N_i are the indicated self-adjoint operators, of the special form $d\Gamma(A_i)$, on the free electron state vector space; the symbol $\sim$ superscribed on an operator indicates its closure (i.e. the extension of the operator to the maximal domain on which it naturally operates), which automatically exists in view of the commuting and self-adjoint character of the R_i and the N_i; while a sum of closed operators is defined as having in its domain all vectors for which the corresponding series of vectors is convergent and to have the obvious value on this domain. At this point we have of course no assurance that there exist any non-zero vectors in this domain.

Having made a rigorous definition of $H_I(t)$, the real question arises, of what relevant results can be obtained regarding it, or the associated field motion. The "finiteness" and "uniqueness" of $H_I(t)$ are of course in question, but to see what "finiteness" should mean here, the role of the interaction hamiltonian in the general theory needs to be considered. In the first instance, what is empirically relevant is the so-called S- (or scattering) operator, which gives the total motion due to the interaction, acting from time $-\infty$ to time $+\infty$, relative to the free motion of the fields without interaction. The matrix element (Sx, y) of this operator determines the transition probability from the state represented by the vector x to that represented by the vector y.

The scattering operator is definable in a formal way as that unitary operator in the representation determined by the physical vacuum, which induces the automorphism $\lim_{t\to\infty} \theta(t)\theta_0(t)^{-1}$, of the algebra of field observables, with the notation used previously, i.e. $\theta(\cdot)$ is the one-parameter group of automorphisms giving the actual temporal development of the field (in the "Heisenberg picture"), while $\theta_0(\cdot)$ gives the kinematics. The precise existence of this limit is of course open to question, but not a great deal more so than the rigorous existence of the transformations $\theta(g)$. It is not difficult to infer from this definition together with the definition of $H_I(t)$, that formally S is the product integral from time $-\infty$ to $+\infty$, of the exponentials of the $H_I(t)$ (the order of the factors from right to left being the temporal order):

$$S = \lim \prod_{k=1}^{n} \exp\,[iH_I(t_k)(t_{k+1} - t_k)],$$

the limit being taken as $t_1 \to -\infty$, $t_n \to +\infty$, and $\max_k |t_{k+1} - t_k| \to 0$. (The variant of this formula in which the exponentials are replaced by the first two terms in the power series expansion was first explicitly presented by Dyson. This approach leads to an intricate power series expansion for S which has finite coefficients only after "renormalization"; is very possibly divergent everywhere even after renormalization; and seems mathematically unpromising. We deal here only with the indicated exponentials.) From this formula for S and the commutativity of the fields at equal times in the case of a "local" theory it is readily inferred that S commutes with all Lorentz transformations, for a relativistic theory.

In the preceding formula the $H_I(t)$ were formulated in the representation determined by the physical vacuum, but if we wish only to obtain the automorphisms representing the field's temporal development, the key feature for the $H_I(t)$ is that the indicated exponentials $\exp [iH_I(t_k)(t_{k+1} - t_k)]$ should exist and be well-defined. This is implied by, and in the present context is virtually equivalent to, the $H_I(t)$ being diagonalizable in a unique fashion, or in mathematical terms, being "essentially self-adjoint." This is much more than requiring that it be densely defined and hermitian, which in turn is much more than appears to be the case for the interaction hamiltonian in practice. However, the classical approach implicitly restricts the representation in which H_I is examined to be one in which the free-field hamiltonian is also finite, a technical requirement by no means physically indicated, and in fact when the role of the representation is understood is seen to be rather contrary to the spirit of renormalization.

The question next arises of whether there exists a representation in which $H_I(t)$ is essentially self-adjoint. What can be proved is that there is always a representation of the photon field canonical variables, which however is dependent on the fermion field variables, for which $H_I(t)$ is essentially self-adjoint. In terms of canonical pairs of photon field variables $P_1, Q_1, P_2, Q_2, \cdots, H_I(t)$ has the form $\sum_{k=1}^{\infty} (P_k A_k + Q_k B_k)^{\sim}$, the A_k and B_k being mutually commutative time-dependent self-adjoint operators which are functions of the fermion field, and so commute with the P's and Q's. This is the same as the interaction hamiltonian for the van Hove model discussed earlier, except that the A_k and B_k are not numbers, but operators. As in that case, there will exist another set of canonical variables $P_1', Q_1', P_2', Q_2', \cdots$ which are functions of the given operators (which may be taken in a free-field representation), and such that the operator $\sum_k (P_k' A_k + Q_k' B_k)^{\sim}$ is essentially self-adjoint. Thus the interaction hamiltonian $H_I(t)$ is finite in this representation in this sense.

This may possibly be a first step in the desired direction, but there remain a number of serious difficulties. These are notably the time-dependence of the representation, its possible lack of uniqueness, and its applicability only to the photon variables. It is not known whether the time-dependence can be removed, the problem arising from the circumstance that, of course, the various $A_k(t)$ and $B_k(t)$ will not commute for different times t. However, the $A_k(t)$ and $B_k(t)$ are smooth functions of t, and it may fairly plausibly be conjectured that over any finite t-interval, $-T < t < T$, there exists a fixed representation in which all $H_I(t)$ are essentially self-adjoint, and that in fact the product integral which defines the unitary transformation taking the field at time $-T$ into that at time T, exists. On the other hand, it is plausible also that over the infinite interval $-\infty < t < \infty$, there exists no such representation. This latter feature might well not be very distressing from a primarily physical standpoint, since what is observed corresponds to the case of a very large T rather than to the case $T = \infty$. But there would remain in any event an apparent uniqueness problem. It is moderately plausible to expect that, in view of the representation independent results described earlier, as well as the fact that in perturbative conventional renormalization theory the representation does not explicitly intervene, that the final results would be independent of the representation, as long as it yields a finite unitary transformation for the implementation of the motion from time $-T$ to time T, and is reasonably smooth analytically, but this is far from having been proved. The problem of dealing with the fermion representation, concerning which there is at least comparable darkness, is relevant to the particle interpretation of the final quantum field state vector space.

The foregoing approach to a treatment which may be convergent in the interaction representation is not only quite speculative, it is much more complicated and intricate than the simple conceptual ideas behind quantum electrodynamics would suggest are appropriate at the foundational level, as appear to be all rigorous approaches which have been advanced. One might, however, hope that in spite of appearances, the theory is really convergent in most naîve form, in which all the operators are represented in the free-field representation. As noted earlier, it is virtually impossible to prove in a mathematically rigorous way that this cannot be the case, since the possibility of an alternative formulation that is physically acceptable cannot be rigorously ruled out. However, the existence of a partial exact solution for the infinitesimal motion makes possible a definite mathematical result whose physical interpretation is fairly inescapable, that the theory is divergent relative to this comparatively restrictive definition of convergence.

The infinitesimal motion from time t to time $t + dt$ will take a dynamical variable X at time t into the dynamical variable given in a formal way by the expression

$$\exp\,[iH_I(t)dt]X\exp\,[-iH_I(t)dt].$$

If X is one of the canonical P's or Q's, the foregoing expression is readily evaluated in closed form, the result being that the following transformation on the P's and Q's is effected:

$$P_k \to P_k + B_k dt, \qquad Q_k \to Q_k - A_k dt.$$

For any finite numerical value of dt it is a material mathematical question whether there exists a unitary transformation effecting this transformation. The theorems cited earlier do not apply directly, because of the operator rather than scalar character of the A_k and B_k, but the same methods apply and indicate in fact that no such unitary operator can exist. Roughly speaking, the A_k and B_k must tend to zero fairly rapidly for such a transformation to be unitarily implementable, and they are actually however of roughly the same order of magnitude for all k. Thereby the possibility of evaluating in closed form the infinitesimal motion of the photon field, in the interaction representation, for quantum electrodynamics "in a box," shows the essential divergence of the latter theory, relative to the most natural conventional requirements. We are led back to the necessity of considering a variety of representations for the field variables.

To summarize, it may be said in general terms that the foregoing approach has yielded quite a small return on a relatively substantial investment. One might be tempted to try an altogether different approach. Now approaches may be classified on the basis of the "picture" they employ—Heisenberg or interaction, to exhaust the apparently reasonable relativistic possibilities—and on the basis of the extent to which the figurative operators of the theory are required to resemble bona fide operators in Hilbert space. The approach indicated above, for example, used the interaction picture, and more significantly, involved a medium level of resemblance to operators; some but by no means all of the pseudo-operators in the formalism are assumed to be rigorously representable by operators in the representation determined by the vacuum state, the other pseudo-operators being representable as operators in other representations, but not necessarily in this one. It is natural to ask whether progress can be made by using instead the Heisenberg picture and assuming the least possible concerning the figurative operators involved.

Perhaps the simplest approach along these lines is that based on the vacuum expectation values of simple products of (Heisenberg) field operators. While the so-called "time-ordered" product expectation values are closer to practice in renormalization theory, they present a more complicated mathematical problem. One may start from the notion that although the interacting field operator $\phi(x)$ does not exist as a mathematical operator, nor possibly even the suitably averaged-out field $\int\phi(x)f(x)d_4x$, with smooth weight function f vanishing outside a sphere, nevertheless, the vacuum expectation values of products of the latter pseudo-operators may have a precise mathematical meaning; it is arguable that these vacuum expectation values are relatively close to measurement. At any rate, in conventional theory, these vacuum expectation values appear as finite after renormalization, although originally divergent; these preliminary infinities might give one pause, but represent almost a positive virtue within the framework of renormalization philosophy for it is presumably the renormalized fields which are physical, and hence finite, while the unrenormalized, "bare" fields are considered virtual objects which might well be infinite from a real physical viewpoint and whose intrusion into the theory one would like to minimize. The "bare" fields are needed at the present time if one wants to construct a specific theory like quantum electrodynamics, but it seems natural to undertake to develop methods for going as far as possible with the use of the more physical renormalized fields.

There are, however, a number of foundational difficulties with this general approach. In order to arrive at non-negative transition probabilities, it is necessary to assume that the vacuum expectation values satisfy positive definiteness conditions, such as are automatically satisfied if they are actually expectation values of operators. These expectation values are in fact quite analogous to the moments of a probability distribution (albeit in an infinite-dimensional space and with non-commuting random variables), and the analogues to familiar non-negativity requirements on certain functions of these moments must hold. Unfortunately, it is not known whether these positive definiteness conditions continue to hold after renormalization. In the absence of any direct formal argument to this effect, it would provide some reassurance to know that it holds in perturbation theory in some non-trivial special case, but as yet there is no result in this direction, as is not so surprising, the computations involved being quite formidable. On the whole, however, some skepticism regarding the positive definiteness of the vacuum expectation values of the renormalized fields, in quantum electrodynamics for example, seems justified.

On the other hand, it may be argued in a semi-empirical way that the positive definiteness conditions must be satisfied, inasmuch as real transition probabilities are non-negative. However, if this is the case, then the vacuum expectation values make it possible to set up a Hilbert space, and a concrete representation of the field operators on this Hilbert space along lines similar to those indicated above for a state (expectation value functional) on a C^*-algebra. The Lorentz-invariance of the vacuum expectation values would lead to a unitary representation of the Lorentz group on this Hilbert space, etc. If one begins to assume that the resulting field operators have reasonable regularity properties, e.g. that $\int\phi(x)f(x)d_4x$ is represented by an essentially self-adjoint operator, then one has not gained much by concentrating on vacuum expectation value functionals rather than operators. Such regularity properties are required in any event for uniqueness of the probability distributions for the pseudo-observables of the theory.

It is also mathematically entirely possible to have a well-defined regular state in which the vacuum expectation values of products of field operators are infinite. Another possibility is that the vacuum expectation values of the products are all finite, but fail to uniquely determine the state, just as the moments of a distribution do not necessarily uniquely determine the distribution. Still another difficulty could arise with finite vacuum expectation values for products of field operators determining a unique regular state, which, however, fail to satisfy the stringent continuity requirements on these values as functions of the weight functions f, which are involved in the effective analytical development of the theory—as happens even in the case of certain free fields of vanishing mass.

There are many other formulations which might be considered, some of which have been intensively developed, but the foregoing should give a fair idea of the complex and novel mathematical problems present, in rather varying respects, in all of them. These problems do not represent grounds for undue pessimism, and we believe that the representation-independent operator-theoretic approach indicated above is relatively promising, but it is unrealistic to anticipate a simple solution to the problem of formulating conventional quantum electrodynamics, say, in a rigorous mathematical manner in the near future. In fact, there may never be a conceptually simple and rigorous way of treating conventional quantum field theory, for this involves analyzing the states of the field in terms of the states of a free field, which is physically somewhat mythical, especially from the point of view of renormalization theory. Easy empirical applicability results from the way in which "free physical particles" may be brought into the theory, but not only may logical consistency be sacrified thereby, conceivably the possibility of a fundamentally simple and rigorous mathematical theory may be eliminated by what may be, from the point of view of empirically applied mathematics, a brilliant approximation. In the next chapter we shall briefly indicate an approach to the theory of interacting fields which is more intrinsic—independent of any ad hoc linear reference system, or Lagrangian, or hamiltonian—and apparently more conceptually satisfactory, although much further from dealing in an effective computational way at this time with real empirical effects.

References to Chapter VII

A number of books on renormalized quantum electrodynamics are available among which we cite that of Källén (1958) for an account emphasizing the Heisenberg picture and that of Jauch and Rohrlich (1955) for an account emphasizing the interaction picture.

Concerning vacuum expectation values of products of fields, see especially Wightman (1956) and later articles by the same author.

The theorem quoted concerning the finiteness of linear forms in canonical Bose-Einstein field variables, in a suitable representation, is proved by Segal (1960a).

CHAPTER VIII

New Approaches and Problems

As indicated earlier, none of the existing approaches to a rationalization and rigorization of conventional quantum field theory has yet progressed to a point within sight of comprehensive definitive results; and even if such progress were at hand, there would remain an unsatisfactory element in the dependence of the theory on an ad hoc linear reference frame. To arrive at a way of treating in an intrinsic fashion the quantization of a non-linear system, it seems reasonable to begin with the analogous finite-dimensional problem (which was mentioned earlier in these chapters). Further, since our primary aim is a new formalism, it is reasonable to proceed initially in a partially heuristic way, from a mathematical standpoint, if this appears to promote succinctness and clarity for the essential formal elements.

Let then $\mathscr{S}$ denote a finite-dimensional smooth manifold which is the configuration space of some system. When $\mathscr{S}$ is a linear vector space $\mathscr{L}$, it may be quantized by forming the phase space $\mathscr{M} = \mathscr{L} \oplus \mathscr{L}^*$, on which there will be canonically defined, as noted earlier, a skew form B. There then exists an essentially unique map $z \to R(z)$ from $\mathscr{M}$ to the self-adjoint operators in some Hilbert space, which is suitably linear, and such that the Weyl relations, which in their infinitesimal form assert that

$$[R(z), R(z')] \sim -iB(z, z'),$$

are satisfied, for arbitrary elements z and z' of $\mathscr{M}$. The analogous object to $\mathscr{M}$ for a general manifold is the so-called cotangent bundle, which consists of all pairs (q, p), where q is a point of $\mathscr{S}$ and p is in the dual to the tangent space to $\mathscr{S}$ at p (i.e. a so-called covector). Denoting this space by $\mathscr{M}$, there will be in place of the distinguished skew-symmetric bilinear form B for the linear case, a skew-symmetric bilinear form Ω_z in the tangent vectors to M at the point z. This form, as a function of z, is but one way of specifying a second-order differential form Ω on M, which may in fact be given by the classical formula

$$\Omega = \sum_k dp_k dq_k,$$

where the $q_1, \cdots, q_n$ are arbitrary local coordinates at a point of $\mathscr{S}$, and $p_1, \cdots, p_n$ are the corresponding covector coordinates.

Now Ω cannot be used in quite as simple a fashion as B in the linear case to set up commutation relations. The generalized canonical variables entering into such relations should on the one hand relate to the form Ω, and on the other specialize in the linear case to essentially the usual variables. It is natural to bring in in place of the vector z in the linear phase space a vector field on the general phase space $\mathscr{M}$, for these relate to Ω in somewhat the same manner as the z's relate to B, and the vector z determines a vector field in a canonical way, namely that generating displacement through a vector proportional to z. Such are of course extremely special vector fields so that it might appear that associating a canonical variable with a general vector field would lead to an excess of such variables, but it turns out in the end that we get no more than are appropriate.

The vector fields Z on $\mathscr{M}$ will generally not commute, unlike the translations in a linear space, so that some additional complication such as that represented by a further term in the commutation relations might be anticipated. In this way one is led to the generalized commutation relations

$$-i[R(Z),R(Z')] = \Omega(Z,Z') + R([Z,Z']).$$

It may be seen that these extend not only the defining relations for $R(z)$ for the case when $\mathscr{S}$ is a linear space but also the commutation relations between the usual angular momenta, linear momenta, and position coordinates in quantum mechanics, constituting in fact a formulation of all such relations invariant under all classical contact transformations.

Such a transformation T on the phase space $\mathscr{M}$ is one leaving invariant the fundamental form Ω. It will naturally leave invariant also every power of Ω, in particular Ω^n for n equal to the dimension of S. In classical terms, this form may be expressed by the equation

$$\Omega^n = \prod_k dp_k\, dq_k,$$

so it follows that any contact transformation leaves invariant this element of measure on $\mathscr{M}$. It follows that the corresponding transformation on the Hilbert space $\mathscr{H} = L_2(M, \Omega^n)$ of all square-integrable functions on $\mathcal{M}$ with respect to this measure,

$$f(z) \to f(T^{-1}z),$$

is unitary, an observation first made by Koopman, and exploited by him and others in connection with classical mechanics. The space $\mathcal{H}$ is useful at this point primarily in affording a convenient representation for the generalized canonical commutation relations just described. Specifically, let ω denote the differential form $\sum_k p_k\, dq_k$ on $\mathcal{M}$, an invariant under general transformations on $\mathcal{S}$, or rather, their canonically induced action on $\mathcal{M}$. Evidently, the covariant differential $d\omega$ is the original form Ω, in the sense of the theory of exterior differential forms, and setting

$$R(X) = -iX + \frac{1}{2}\omega(X),$$

it is straightforward to verify that the generalized commutation relations are satisfied and that $R(X)$ is hermitian when X is an infinitesimal contact transformation.

It can now be observed that it does not matter that the space $\mathcal{M}$ and form Ω are associated with a configuration space $\mathcal{S}$. The foregoing quantization depends only on the phase space structure. If there is given only a manifold $\mathcal{M}$ representing the phase space of some physical system, without any special labelling of the local coordinates as "spatial" on the one hand or "momentum-like" on the other, together with a distinguished non-degenerate second-order differential form Ω on $\mathcal{M}$, which is closed: $d\Omega = 0$, then similar commutation relations may be set up without inconsistency. In case $\mathcal{M}$ is simply-connected, there will exist a form ω on $\mathcal{M}$ such that $\Omega = d\omega$, and a construction entirely similar to that above is possible. The resulting set of canonical variables is essentially unique, different choices for ω giving rise to unitarily equivalent constructions. The expendability in this connection of an actual configuration space $\mathcal{S}$ is noteworthy because in a relativistic field theory there is no fully covariant way to distinguish the p's from the q's, or to regard the phase space as the cotangent bundle of some distinguished manifold $\mathcal{S}$; a useful non-relativistic $\mathcal{S}$ may be defined, but the full Lorentz group will not act on $\mathcal{S}$.

The case of an infinite-dimensional manifold $\mathcal{M}$ of the type that occurs in quantum field theory is the one of real concern here; let us consider now the adaptation of the foregoing approach to the quantization problem for such a manifold. The simplest representative case is apparently the manifold $\mathcal{M}$ of all solutions of a Lorentz-invariant partial differential equation such as

$$\Box\phi = F(\phi),$$

where F is a smooth function of a real variable such that $F(0) = 0$ and $F'(\lambda) \geqq 0$ for all real values of λ, the unknown function ϕ being real-valued. (The last condition replaces the reality of the mass in the Klein-Gordon equation.) The manifold $\mathcal{M}$ of all solutions ϕ of this differential equation is, as a point set, not a particularly accessible object at the present stage of development of the global theory of non-linear partial differential equations, but the point set aspect is secondary. Primarily, in relation to quantization, $\mathcal{M}$ is important as a variety of measure space on which the Lorentz group acts, as is the Hilbert space for the renormalized Schrödinger representation, and like the Hilbert space, may contain sets which are large by topological standards, but are effectively of measure zero, and do not affect the quantization. Further, despite the relative inaccessibility of $\mathcal{M}$, its tangent manifolds are defined by simple linear partial differential equations: the tangent space at a point ϕ may be parametrized by functions η satisfying the first-order variation of the given equation:

$$\Box\eta = F'(\phi)\eta.$$

This is a linear hyperbolic equation with constant principal part, and despite relevant gaps in the global spectral theory for such equations, there are a good many recent results and methods pertinent to them. One way to look at the manifold $\mathcal{M}$ is to observe that the indicated tangent plane T_ϕ may be defined by the foregoing equation for an arbitrary ϕ, whether a solution of the given non-linear equation or not; $\mathcal{M}$ may then be regarded as a maximal integral manifold in general function space of the elements of contact given by the correspondence $\phi \to T_\phi$, namely that passing through the point $\phi = 0$. The usual integrability conditions are automatically satisfied by virtue of the manner of construction of the elements of contact. In any event, while the analytical difficulties involved in the construction of $\mathcal{M}$ must not be minimized, the problem seems quite an approachable one.

The application of the method indicated earlier for finite-dimensional manifolds to the present infinite-dimensional one depends on associating with $\mathcal{M}$ a suitable form Ω (i.e. imposing on $\mathcal{M}$ a so-called distinguished "symplectic" structure). The fundamental differential form Ω may in fact be defined in the following way. First, the "commutator function" $D_\phi(x, x')$ at the point ϕ of $\mathcal{M}$ may be defined as that solution of the hyperbolic equation

$$\Box\, D(x) = F'(\phi)D,$$

which has the Cauchy data

$$D_\phi(x, x')\Big]_{x_0 = x_0'} = 0, \qquad \frac{\partial}{\partial x_0} D_\phi(x, x')\Big]_{x_0 = x_0'} = \delta(\mathbf{x} - \mathbf{x}').$$

Next, to avoid delicate and secondary real variable questions, let the tangent plane T_ϕ at ϕ, which was set up originally as the null space of the linear operator $\Box - F'(\phi)$ be formulated rather as the quotient space modulo the range of this operator, considered, for definiteness, as acting on the space of infinitely differentiable functions of compact support. In view of the hermitian character of the operator in question, these two spaces are formally identifiable. For simplicity, ϕ and F may also be assumed infinitely differentiable. If η and η' are any two tangent vectors, let f and f' be any representatives in the residue classes defining them, and define

$$\Omega_\phi(\eta, \eta') = \int\!\!\int D_\phi(x, x')f(x)f(x')\, d_4x\, d_4x',$$

giving the differential form Ω on $\mathcal{M}$. (Alternatively, we may consider only tangent vectors η of the form

$$\eta(x) = \int D_\phi(x, x')f(x')\, d_4x',$$

and use the same definition for Ω_ϕ.)

That Ω is closed can be derived by a formal argument. It is then reasonable to assume that canonical variables satisfying the indicated commutation relations exist. A formal quantum field ϕ satisfying the conventional commutation relations can then be defined in the following way. For any smooth function f on space-time, let X_f be the vector field on $\mathcal{M}$ which assigns at any point ϕ of $\mathcal{M}$ the tangent vector in the residue class (modulo the range of $\Box - F'(\phi)$) determined by f. Since X_f depends linearly on f, and $R(X)$ depends linearly on X, $R(X_f)$ depends linearly on f, and so is formally expressible as

$$R(X_f) \sim \int \boldsymbol{\phi}(x)f(x)\, d_4x,$$

for some operator-valued pseudo-function $\boldsymbol{\phi}$. The commutation relations for the $R(X)$ given above then carry corresponding implications for the $\phi(x)$, and it may be seen in this way that in particular the following relations are satisfied:

$$[\phi(x), \phi(x')] = 0, \qquad \left[\phi(x), \frac{\partial\phi(x')}{\partial x_0'}\right] = \delta(\mathbf{x} - \mathbf{x}') \quad \text{for } x_0 = x_0'.$$

Thus the operator field $\phi(x)$ satisfies the conventional canonical commutation relations, and being canonically derived from the classical equations of motion, is reasonably definable as the quantized field associated with the classical system defined by the given partial differential equation.

The basic algebra of operators is thereby formally established. The question next arises of the determination of the vacuum state. It is reasonable, physically and mathematically, to define this as a regular Lorentz-invariant state (in the sense of expectation value functional), in the representation determined by which the energy is non-negative. It is plausible, as is known rigorously in the case of free fields, that the vacuum is then unique. The vacuum state then determines a representation for the generalized canonical variables $R(X_f)$ with a vacuum state representative as cyclic vector, as well as a unitary representation of the Lorentz group, on a Hilbert space $\mathscr{K}$, the space of physical quantum field state vectors. There is a price to be paid for this formalism—it is difficult to make a simple particle interpretation for the states of $\mathscr{K}$. But the general ambiguity in the physical notion of particle indicates that this is just as it should be. The existence of a simple particle interpretation for all the states of a quantum field must be expected to be quite a special property of the interacting field equations, and there are in fact formal possibilities in this direction for suitable equations.

The foregoing was intended to sketch a possible framework for quantum field theory which avoids the conventional dependence on a linear reference system, and at the same time is free of the conventional divergences, at least as long as a particle interpretation of the quantum field is not enforced. It was intended at the same time to illustrate how modern mathematical ideas may be brought to bear on theoretical physical problems, as well as how mathematical problems can grow out of general physical considerations. It is clear that to follow up mathematically the heuristic developments indicated requires the parallel development of the global theory of non-linear hyperbolic equations on the one hand, and analysis and aspects of differential geometry on infinite-dimensional smooth symplectic manifolds on the other. Further thought leads to relatively specific problems along these lines. Among these may be cited the Cauchy problem with covariant (time-dependent) data at infinity for variable coefficient and non-linear hyperbolic equations, in the realm of partial differential equations, and that of the generalization of the basic theorem for finite-dimensional manifolds on the existence of manifolds with prescribed elements of contact, in the light of spectral theory in Hilbert space, in the realm of analytical differential geometry in infinite-dimensional manifolds. Also relevant although less obviously so are the global analytical properties of non-linear transformations in Hilbert space. The cited Cauchy problem provides a canonical formal coordinatization of $\mathcal{M}$ by the tangent plane T_0 at the vanishing field.

We shall conclude by discussing quite a different sort of problem, which likewise illustrates the interplay between mathematics and theoretical physics, and is of fundamental importance for modern physics. The problem of quantum field theory is primarily one of meaning, especially of a mathematical sort; there is as yet no rigorous proof that renormalization is not in essence some extremely sophisticated variety of witchcraft. But despite this problem of meaning, of the product of local fields in the mathematical and physical senses, etc., the numbers produced fit very nicely with experiment. The general situation is the reverse in the theory of the classification of free elementary particles. It is fairly well understood in a theoretical way how particles may be classified in relation to a given symmetry group, and there are no serious convergence questions or problems of mathematical meaning. But the classification schemes produced theoretically have very limited power for empirical prediction, and tend to be quite short-lived. In a formal way, the discussion up to this point treated the quantization of a given non-linear manifold $\mathcal{M}$; the question now is, what can be said about this manifold, specifically, about the structure of the tangent space to it at the point representing the vanishing field (or more generally, the presumably unique point left invariant by the fundamental symmetry group)? This tangent space can be regarded as the space of wave functions for the free quanta of the theory, whose interaction is determined by the differential-geometric structure of $\mathcal{M}$.

Although ever since the foundation of modern quantum theory around 1925, leading physicists, such as Niels Bohr, have emphasized the need for radically new notions of space and time, insofar as microphenomena are concerned, virtually all highly developed classification schemes for particles or for reduction of experimental data have been based on Minkowski space, and the action of the Lorentz group or even in some cases on non-relativistic models, up to the present time. The time would appear to be much more than ripe for the development of more imaginative, sophisticated, and rational, theoretical schemes for classification. There are, however, two notable difficulties about this. First of all it is obvious that only a quite comprehensive and detailed scheme, with an accompanying clearcut physical interpretation has any chance of dealing with the complicated phenomena that have been observed. Secondly, at first glance it might appear that one might suffer from an embarrassment of riches in the way of possible classification schemes. But a closer look at the situation indicates that these difficulties may not be so substantial.

The two most fundamental advances in mechanics since the time of Newton may both be formulated as cases in which a certain Lie algebra of mechanical significance is replaced by a less degenerate algebra. The passage from classical mechanics to quantum mechanics involved above all the introduction of the relation $[p, q] = -i\hbar$, which as $\hbar \to 0$ degenerates into the commutativity characteristic of classical mechanics. Similarly relativistic mechanics passes into non-relativistic mechanics as $c \to \infty$, corresponding to the degeneration of the Lorentz group into the Galilean group. In precise mathematical terms what happens is that the constants of structure of the Lie algebra of the Lorentz group, which are c-dependent, converge as $c \to \infty$ to those of the Galilean group. It is natural to inquire as to whether it is possible for the basic variables of the conventional relativistic theory to be a degenerate form of those of a more accurate and effective theory, and if so, what possibilities there may be.

The basic variables of the standard relativistic theory are 14 in number: the linear momenta and energy, p_0, p_1, p_2, p_3; the angular momenta, m_1, m_2, m_3; the space-time coordinates x_0, x_1, x_2, x_3; and the parasitic but unavoidable Lorentz momenta, non-measurable in view of their lack of commutation with the energy, but needed for a covariant theory. The commutation relations of the generators of the Lorentz group are well known, while those of the x_k with these generators are implicit in the usage, which requires that any Lorentz transformation transforms the x_k in an inhomogeneous linear style. To arrive at a Lie algebra the identity I must be included, as it arises from the commutator of p_k with x_k, and all together a 15-dimensional algebra is obtained. Now one difficulty of long standing in the standard relativistic theory has been the lack of any bona fide coordinate operators for the standard relativistic particles—the photon, electron, etc. Many authors have examined this question but no fully covariant set of coordinate operators has been found, and it can in fact be proved that there do not exist commuting self-adjoint operators x_0, x_1, x_2, x_3 on the Hilbert space of classical wave functions for any of these particles, which transform in the required fashion under the action of the Lorentz group; or what is equivalent, the standard irreducible unitary representations of the Lorentz group cannot be extended to unitary representations of the 15-dimensional group whose Lie algebra was cited above, without enlargement of the representation space. There are actually many other difficulties with the standard spaces of relativistic wave functions as models for free particles.

Now there exist actually many Lie algebras of which the standard relativistic one is a degenerate form, and three are known in fact which are terminal, in the sense that they are not themselves degenerate forms of any other Lie algebras. These are the Lie algebras of the non-compact real pseudo-orthogonal groups in six variables. There are also well-defined representations of these groups which degenerate into the familiar representations of the Lorentz group discussed above. Associated models have certain empirically suggested qualitative features, and also provide a notably economical possible means of integrating the so-called "internal" and "external" degrees of freedom of particles. A number of current essentially conventional models simply adjoin to the Lorentz group, as the symmetry group relevant to the external degrees of freedom, a separate "internal" group, variously of three and six dimensions, the total effective symmetry group being the direct product group and having so large a dimension as to render a conclusive experimental test of its possible validity extremely difficult. Despite the relative economy of the 15-dimensional groups described, which are roughly of the right size to provide with a small leeway for the "constants of motion" which have been established experimentally, the greatest difficulty with the associated models may well be that the existing data cannot differentiate between them, rather than that they fail to fit the data.

While these groups and certain four-dimensional spaces on which they act seem particularly worth investigating from the point of view indicated, many other symmetry groups and associated representations and physical interpretations may be conceived of. It seems safe to say that a total symmetry group of dimension less than 10 is likely to be much too restrictive to serve as a basis for the description of the known particles, while one of dimension greater than 20 could probably not be conclusively established on the basis of the kind of data about the particles likely to be available in the foreseeable future; but there is little beyond this that is both safe and simple. In any event, after a tentative fundamental symmetry group (or, equivalently, Lie algebra) has been selected, the main steps involved in formulating a specific physical theory of elementary particles may be outlined briefly as follows.

First, certain linear representations of the group must be specified and connected with designated elementary particles (where "elementary" does not necessarily have any absolute meaning, but refers only to its empirically observed physical role). Second, a maximal abelian diagonalizable subalgebra of the group algebra must be designated; the spectral values for the elements of the subalgebra give the so-called "quantum numbers" for the particles in question. Usually it is the infinitesimal group algebra, or so-called enveloping algebra of the Lie algebra, which is employed, after augmentation by the quite limited subgroup of elements in the absolute center of the group, involving only one non-trivial element in the relativistic case, which specifies whether the spin is integral or half integral. Thirdly, these quantum numbers must be connected with experimentally measurable quantities, which involves incidentally the construction of a dictionary between the quantum numbers and conventional ones employed in connection with the standard relativistic theory as augmented by various internal quantum numbers, such as strangeness, baryon number, etc.

Thus it is a long road to the formulation of a complete theory, but it is a finite one. When it is finished it may be checked with experiment. Despite the large number of conceivable models, there are relatively few of outstanding economy and other desirable qualitative features. The development of any such models is a rather substantial undertaking, but it seems as sensible to make a controlled, physically conservative (if theoretically imaginative) attempt such as very briefly indicated as to pursue further variations on classification schemes via the Lorentz group and various uncertain "internal" symmetry groups.

Even from a purely mathematical point of view, the determination of the relevant representation theory of the groups mentioned, and the establishment of an effective theory of the convergence of group representations, is an interesting challenge.

On the other hand, there is a definite risk in any attempt at dealing with the classification of *free* particles, that the interaction effects are conceivably so great as to completely dominate the empirical masses, coupling constants, etc. The theory may be so strongly non-linear that no linear approximation to it has any recognizable connection with experiment. This is true just as it is true that if the sun had been of small mass, we would never have heard of Kepler. But it does not look particularly likely, in view of the excellent description of the kinematics of the best known particles provided by linear equations such as the Maxwell and Dirac equations over a rather wide energy range.

References to Chapter VIII

The indicated approach to the quantization of non-linear partial differential equations is due to Segal (1960b). The general theory of linear hyperbolic partial differential equations has been treated at length by Leray (1953). Relatively simple conditions for the absolute continuity of a transformation in Hilbert space have been obtained by Gross (1960). Degeneration of one Lie algebra into an inequivalent one, in relation to the classification of elementary particles, has been considered by Segal (1951, 1959a). Recently K. Jörgens (1961) has treated the global Cauchy problem for a class of non-linear relativistic partial differential equations.

Bibliography

G. Birkhoff and J. von Neumann

1936 *The logic of quantum mechanics.* Ann. of Math. (2) **37**, 823–843.

N. Bohr and L. Rosenfeld

1933 *Zur Frage der Messbarkeit der Elektromagnetischen Feldgrössen.* Mat.-Fys. Medd. Danske Vid. Selsk. **12**, no. 8 (65 pp.).

R. Brauer and H. Weyl

1935 *Spinors in n dimensions.* Amer. J. Math. **57**, 425–449.

R. H. Cameron and W. T. Martin

1945 *Fourier-Wiener transforms of analytic functionals.* Duke Math. J. **12**, 489–507.

1947 *The behavior of measure and measurability under change of scale in Wiener space.* Bull. Amer. Math. Soc. **53**, 130–137.

1949 *The transformation of Wiener integrals by nonlinear transformations.* Trans. Amer. Math. Soc. **66**, 253–283.

J. M. Cook

1953 *The mathematics of second quantization.* Trans. Amer. Math. Soc. **74**, 222–245.

P. A. M. Dirac

1927 *The quantum theory of the emission and absorption of radiation.* Proc. Roy. Soc. London Ser. A **114**, 243–265.

1947 *The principles of quantum mechanics.* 4th ed., Oxford, 1948.

F. J. Dyson

1949a *The radiation theories of Tomonaga, Schwinger, and Feynman.* Phys. Rev. (2) **75**, 486–502.

1949b *The S-matrix in quantum electrodynamics.* Phys. Rev. (2) **75**, 1736–1755.

R. P. Feynman

1950 *Mathematical formulation of the quantum theory of electromagnetic interaction.* Phys. Rev. (2) **80**, 440–457.

1951 *An operator calculus having applications in quantum electrodynamics.* Phys. Rev. (2) **84**, 108–128.

V. Fock
1932 *Konfigurationsraum und zweite Quantelung.* Z. Physik **75**, 622–647; **76**, 952.

K. O. Friedrichs
1953 *Mathematical aspects of the quantum theory of fields.* Interscience, New York. (Parts 1–5 originally published in Comm. Pure Appl. Math. **4–6**, 1951–1953.)

L. Gårding and A. S. Wightman
1954 *Representations of the commutation and anticommutation relations.* Proc. Nat. Acad. Sci. U.S.A. **40**, 617–626.

I. Gelfand and M. Neumark
1943 *On the imbedding of normed rings into the ring of operators in Hilbert space.* Mat. Sb. (N.S.) **12**, 197–213.

L. Gross
1960 *Integration and nonlinear transformations in Hilbert space.* Trans. Amer. Math. Soc. **94**, 404–440.

R. Haag
1955 *On quantum field theories.* Mat.-Fys. Medd. Danske Vid. Selsk. **29**, no. 12 (37 pp.).

W. Heisenberg and W. Pauli
1929/1930 *Zur Quantendynamik der Wellenfelder.* Z. Physik **56**, 1–61; **59**, 168–190.

J. M. Jauch and F. Rohrlich
1955 *The theory of photons and electrons.* Addison-Wesley, Cambridge, Massachusetts, 488 pp.

K. Jörgens
1961 *Das Anfangswertproblem im Grossen für eine Klasse nichtlinearer Wellengleichungen.* Math. Z. **77**, 295–308.

S. Kakutani
1950 *Determination of the spectrum of the flow of the Brownian motion.* Proc. Nat. Acad. Sci. U.S.A. **36**, 319–323.

G. Källén
1958 *Quantenelektrodynamik. Handbuch der Physik* **5**, pt. 1, Berlin, pp. 169–364.

G. Källén and A. S. Wightman
1958 *The analytic properties of the vacuum expectation value of a product of three scalar local fields.* Mat.-Fys. Skr. Danske Vid. Selsk. **1**, no. 6 (58 pp.).

H. Lehmann, K. Symanzik, and W. Zimmermann
1955/1957 *On the formulation of quantized field theories.* Nuovo Cimento (10) **1**, 205–225; **6**, 319–333.

J. LERAY
1953 *Hyperbolic differential equations.* Institute for Advanced Study, Princeton, N.J. 240 pp. (mimeographed).

D. B. LOWDENSLAGER
1957 *On postulates for general quantum mechanics.* Proc. Amer. Math. Soc. **8**, 88–91.

G. W. MACKEY
1949 *A theorem of Stone and von Neumann.* Duke Math. J. **16**, 313–326.
1957 *Quantum mechanics and Hilbert space.* Amer. Math. Monthly **64**, 45–57.
1958 *Unitary representations of group extensions.* I. Acta Math. **99**, 265–311.

F. J. MURRAY and J. VON NEUMANN
1936 *On rings of operators.* Ann. of Math. (2) **37**, 116–229.

J. VON NEUMANN
1927 *Mathematische Begrundung der Quantenmechanik.* Göttinger Nachr. 1–57. *Wahrscheinlichskeitstheoretischer Aufbau der Quantenmechanik*, ibid. 245–272.
1931 *Die Eindeutigkeit der Schrödingerschen Operatoren.* Math. Ann. **104**, 570–578.
1932 *Mathematische Grundlagen der Quantenmechanik.* Berlin. (Available also in English translation.)
1936 *On an algebraic generalization of the quantum mechanical formalism.* Mat. Sb. (N.S.) **1**, 415–484.
1938 *On infinite products.* Compositio Math. **6**, 1–77.

S. S. SCHWEBER and A. S. WIGHTMAN
1955 *Configuration space methods in relativistic quantum field theory.* I. Phys. Rev. (2) **98**, 812–837.

J. SCHWINGER
1949 *Quantum electrodynamics.* I, II, III. Phys. Rev. (2) **74**, 1439–1461; **75**, 651–679; **76**, 790–817.
1951 *On the Green's functions of quantized fields.* Proc. Nat. Acad. Sci. U.S.A. **37**, 452–459.
1951/1953 *The theory of quantized fields.* I, II, III. Phys. Rev. (2) **82**, 914–927; **91**, 713–728; **91**, 728–740.

I. E. SEGAL
1947a *Irreducible representations of operator algebras.* Bull. Amer. Math. Soc. **53**, 73–88.
1947b *Postulates for general quantum mechanics.* Ann. of Math. (2) **48**, 930–948.

1951 *A class of operator algebras which are determined by groups.* Duke Math. J. **18**, 221–265.

1956a *Tensor algebras over Hilbert spaces.* I. Trans. Amer. Math. Soc. **81**, 106–134.

1956b *Tensor algebras over Hilbert spaces.* II. Ann. of Math. (2) **63**, 160–175.

1957 *The structure of a class of representations of the unitary group on a Hilbert space.* Proc. Amer. Math. Soc. **8**, 197–203.

1958 *Distributions in Hilbert space and canonical systems of operators.* Trans. Amer. Math. Soc. **88**, 12–41.

1959a *Caractérisation mathématique des observables en théorie quantique des champs et ses conséquences pour la structure des particules libres.* Colloque internationale sur les problèmes mathématiques de la théorie quantique des champs. Paris, publ. by C.N.R.S., pp. 57–103.

1959b *Foundations of the theory of dynamical systems of infinitely many degrees of freedom.* I. Mat.-Fys. Medd. Dansk. Vid. Selsk. **31**, no. 12 (38 pp.).

1960a *Quasi-finiteness of the interaction hamiltonian of certain quantum fields.* Ann. of Math. (2) **72**, 594–602.

1960b *Quantization of nonlinear systems.* J. Math. Phys. **1**, 468–488.

1961 *Foundations . . .* II. *A generating functional for the states of a linear boson field.* Canad. J. Math. **13**, 1–18.

1962 *Mathematical characterization of the physical vacuum.* Illinois J. Math. **6**, 500–523.

D. SHALE

Linear symmetries of free boson fields. Trans. Amer. Math. Soc. **103**, 149–167.

S. SHERMAN

1951 *Non-negative observables are squares.* Proc. Amer. Math. Soc. **2**, 31–33.

1956 *On Segal's postulates for general quantum mechanics.* Ann. of Math. (2) **64**, 593–601.

M. H. STONE

1930 *Linear transformations in Hilbert space.* III. *Operational methods and group theory.* Proc. Nat. Acad. Sci. U.S.A. **16**, 172–175.

1932 *On one-parameter unitary groups in Hilbert space.* Ann. of Math. (2) **33**, 643–648.

1940/1941 *A general theory of spectra.* I, II. Proc. Nat. Acad. Sci. U.S.A. **26**, 280–283; **27**, 83–87.

L. VAN HOVE

1952 *Les difficultés de divergence pour un modéle particulier de champ quantifié.* Physica **18**, 145–159.

H. WEYL

1927 *Quantenmechanik und Gruppentheorie.* Z. Physik **46**, 1–46.

N. WIENER

1938 *The homogeneous chaos.* Amer. J. Math. **60**, 897–936.

N. WIENER and A. SIEGEL

1953 *A new form for the statistical postulate of quantum mechanics.* Phys. Rev. (2) **91**, 1551–1560.

A. S. WIGHTMAN

1956 *Quantum field theory in terms of vacuum expectation values.* Phys. Rev. (2) **101**, 860–866.

E. P. WIGNER

1939 *On unitary representations of the inhomogeneous Lorentz group.* Ann. of Math. (2) **40**, 149–204.

Group Representations in Hilbert Space

By G. W. MACKEY

1. Introduction. Though this appendix will be mainly of a purely mathematical character we shall begin with an attempt to explain one aspect of the general relevance of the theory of group representations for quantum mechanics.

Let S be a set whose elements have been put into one-to-one correspondence with the "states" of a physical system. Here we suppose that "state" has been defined in such a way that the state of the system at time $t > t_0$ is uniquely determined by the state at time t_0 and the relevant physical law. Let $U_t(s)$ denote the state at time t when the state at time 0 is s. Then the state at time $t_1 + t_2$ is $U_{t_1}(U_{t_2}(s))$ and is also $U_{t_1+t_2}(s)$. Thus we have the identity $U_{t_1+t_2} = U_{t_1}U_{t_2}$ for the transformations U_t. If we assume that our system is "invertible" in the sense that each U_t maps S onto all of S in a one-to-one manner then we may define $U_{-t} = U_t^{-1}$ and U_0 as the identity and obtain a homomorphism of the additive group R of the real line into the group P_s of all permutations of S. The set S, of course, is always a great deal more than an abstract set. It generally has a topology and other structure as well and this structure is preserved by the U_t. Thus $t \to U_t$ is actually a homomorphism of R into the suitably defined group A_s of all "automorphisms" of S. Insofar as states are described by the rectangular coordinates of points in space and their time derivatives any rigid motion in space will define a permutation V_α of S which will in general belong to A_s. Clearly $V_{\alpha_1\alpha_2} = V_{\alpha_1}V_{\alpha_2}$ and we have a homomorphism $\alpha \to V_\alpha$ of the group G of all rigid motions of space into A_s. Many systems of importance are spatially homogeneous in the sense that $V_\alpha U_t = U_t V_\alpha$ for all t and α. For such we have a homomorphism of the direct product $R \times G$ into A_s. For an important subclass of the spatially homogeneous systems the homomorphism $(t, \alpha) \to V_\alpha U_t$ of $R \times G$ into A_s may be extended to a homomorphism of a larger group of permutations of space time into A_s. For the so-called relativistically invariant systems the domain of the homomorphism may be extended to include the whole inhomogeneous Lorentz group.

In short for many physical systems one has a natural homomorphism of some group G^0 of space time transformations into the group A_s of all automorphisms of the set S of states.

An important structural property of the S for a classical mechanical system (the so-called phase space of the system) is that there is a natural way of assigning an "observable" or "dynamical variable" to each homomorphism of R into A_s. Thus in particular every one-parameter subgroup of G^0 is correlated with a certain observable and the observables arising in this way play a central role in the theory. For example the one-parameter group of translations in time corresponds to the total energy and the one-parameter group of spatial translations in a given direction corresponds to the total linear momentum in that direction. This property persists in quantum mechanics where S is the set of all one-dimensional subspaces of a Hilbert space H_s and so we have a natural way of extending the notions of energy and momentum from classical to quantum mechanics.

In the quantum mechanical case the automorphisms of S are the permutations of the one-dimensional subspaces of H_s which preserve linear independence and orthogonality. It can be shown that every such can be implemented either by a unitary map of H_s on H_s or by an "anti-unitary" map of H_s on H_s (an anti-unitary map is a norm preserving map V of H_s onto H_s which is linear with respect to real scalars and such that $V(i\varphi) = -iV(\varphi)$ for all φ in H_s). For all complex c with $|c| = 1$ it is obvious that cV and V define the same permutation of the one-dimensional subspaces of H_s and hence the same member of A_s. On the other hand it can be shown that this is the extent of the ambiguity. Except for an arbitrary factor cV is uniquely determined by the corresponding member of A_s. Choosing this factor arbitrarily for each $\alpha \in G^0$ we get from our homomorphisms of G^0 into A_s a mapping W, $\alpha \rightarrow W_\alpha$ of G^0 onto the group $U'(H_s)$ of all unitary and anti-unitary operators mapping H_s on H_s. Since $W_{\alpha_1\alpha_2}$ and $W_{\alpha_1}W_{\alpha_2}$ define the same member of A_s we have $W_{\alpha_1\alpha_2} = \sigma(\alpha_1, \alpha_2)W_{\alpha_1}W_{\alpha_2}$ where $\sigma(\alpha_1, \alpha_2)$ is a complex number of modulus one. If the W_α were all unitary and $\sigma(\alpha_1, \alpha_2)$ were identically one $\alpha \rightarrow W_\alpha$ would be what is known as a unitary representation of G^0. When σ is not identically one we have a unitary *ray* or *projective* representation with multiplier σ. Since $W_{\alpha_1\alpha_2} = W_{\alpha_1}W_{\alpha_2}$ is unitary whether W_{α_1} and W_{α_2} are unitary or anti-unitary, it is often possible to prove that no anti-unitaries can arise. In any event we can get rid of anti-unitaries by considering only a certain subgroup of G^0 of index 2. Supposing this done we have associated with every G^0 invariant physical system a certain unitary ray representation of G^0. Any classification of such representations yields a classification of G^0 invariant physical systems and this classification turns out to be significant. In particular when G^0 includes the translations in space and time one can read off from this representation the form of the energy and momentum operators and the functional relationship between energy and momentum. When G^0 includes spatial rotations one obtains information about angular momentum and "spin."

We shall devote the rest of this chapter and the next to an account of a general method for classifying the unitary representations of a class of groups which includes some of those of greatest physical interest. Though this method applies to projective representations we shall, for simplicity's sake, consider only ordinary representations in detail. As an application we shall obtain Wigner's classification of the representations of the inhomogeneous Lorentz group and sketch briefly its significance for the classification of elementary particles.

2. Notation and fundamental concepts. We shall deal with groups G which are separable and locally compact in some topology with respect to which the group operations are continuous. Every such group G admits a non-trivial measure μ defined on all Borel sets and invariant under right translations. This measure is unique up to a multiplicative constant and is called the Haar measure of the group.

Generally speaking a representation of a group is a homomorphism $x \to L_x$ of that group into the group of all non-singular linear transformations of some vector space. We shall deal only with representations in which the non-singular linear transformations L_x are unitary transformations in some separable Hilbert space $H(L)$ and the mapping $x \to L_x$ is strongly continuous in the sense that $L_x(\varphi)$ is a continuous function of x for all φ in $H(L)$. It is useful to note that strong continuity is implied by the superficially much weaker property of weak measurability: $(L_x(\varphi), \psi)$ is measurable as a function of x for all φ and ψ in $H(L)$. Since we shall deal exclusively with strongly continuous unitary representations, there will be no ambiguity in referring to them simply as representations.

Let $L^1, L^2, L^3, \cdots$ be a finite or infinite sequence of representations of G. We define a new representation $L = L^1 \oplus L^2 \oplus \cdots$ known as the *direct sum* of the L^j by taking $H(L)$ to be the direct sum $H(L^1) \oplus H(L^2)$

$\oplus \cdots$ and defining $L_x(\varphi_1, \varphi_2, \cdots)$ as $L_x^1(\varphi_1), L_x^2(\varphi_2), \cdots$. If L and M are two representations of G we define an *intertwining operator* for L and M as a bounded linear transformation T from $H(L)$ to $H(M)$ such that $TL_x = M_xT$ for all $x \in G$. The set of all intertwining operators for L and M is a vector space which we shall denote by $R(L, M)$. If $R(L, M)$ contains a unitary operator V mapping $H(L)$ on $H(M)$, then we may rewrite the defining identity as $M_x = VL_xV^{-1}$. Clearly the representations L and M do not differ in any essential respect and two representations so related are said to be *equivalent*. It is usually not necessary to distinguish between equivalent representations: When $L = M$, then $R(L, M) = R(L, L)$ is a ring of operators called the *commuting ring* of L. If $H(L)$ contains a closed subspace H' which is mapped into itself by all L_x, then the projection on H' is in $R(L, L)$ and conversely. Moreover H'', the orthogonal complement of $H(L)$, will also be mapped into itself by all L_x and L will be equivalent to the direct sum of its restriction to H' and its restriction to H''. These two restrictions are called *subrepresentations* of L. When $H(L)$ contains no non-trivial closed invariant subspaces (or equivalently when L is not equivalent to the direct sum of two non-zero representations) we say that L is *irreducible*. It is not difficult to show that L is irreducible if and only if $R(L, L)$ contains only multiples of the identity. Let T be any non-zero member of $R(L, M)$. Let H_1 denote the orthogonal complement of the null space of T and let H_2 denote the closure of the range of T. Then a modern form of a celebrated lemma of Schur asserts that H_1 and H_2 are invariant subspaces of $H(L)$ and $H(M)$ and define *equivalent* subrepresentations of L and M. It follows in particular that $R(L, M)$ reduces to the zero element if and only if no subrepresentation of L is equivalent to any subrepresentation of M. In this event we shall say that L and M are *disjoint* representations of G.

3. The decomposability of representations. When G is not only locally compact but compact it follows from the celebrated Peter-Weyl theorem that every representation L of G is equivalent to a direct sum $L^1 \oplus L^2 \oplus \cdots$ where the L^j are all irreducible and finite dimensional. Moreover using Schur's lemma it is not hard to establish the following uniqueness theorem: If $L^1 \oplus L^2 \cdots$ is equivalent to $M^1 \oplus M^2 \cdots$ where the L^i and the M^j are irreducible, then there exists a permutation π of $1, 2, \cdots$ such that L^j is equivalent to $M^{\pi(j)}$ for all j. Thus if we know the most general irreducible representation of the compact group G to within equivalence, we know immediately the most general representation. It is uniquely specified by assigning a "multiplicity" $= \infty, 0, 1, 2, \cdots$ in an arbitrary manner to each equivalence class of irreducible representations. For many compact groups the irreducible representations are "known" in a very explicit sense. We cite one example for later use. Let G be the group of all rotations about 0 in Euclidean 3-space E^3. Let S_2 denote the surface of the unit sphere in E^3 and let H denote the Hilbert space of all square-integrable complex valued functions on S_2. For each $j = 0, 1, 2, \cdots$ the subspace of H defined by restricting homogeneous harmonic polynomials of degree j in x, y, z to S_2 is $2j + 1$ dimensional and invariant under the action of G. Thus it defines a representation D^j of G which can be shown to be irreducible. Moreover every irreducible representation of G can be shown to be equivalent to a unique D^j.

For most non-compact groups it is simply not true that every representation is a direct sum of irreducibles and the problem of describing the most general representation in terms of irreducible ones becomes much more difficult when it can be solved at all. It is natural to try to replace the discrete direct sums described above by suitably defined "continuous direct sums" or "direct integrals." Such a notion was defined by von Neumann in [10] and applied by Mautner in [9] to prove that every representation of a separable locally compact G is equivalent to a direct integral of irreducible representations. Unfortunately Mautner's decomposition is highly non-unique. On the other hand for representations which are of "type I" in a sense to be described below Mautner's decomposition is essentially unique and can be made to lead to a quite satisfactory reduction of the general problem to that of finding the irreducible representations. This is especially interesting in view of the fact that many (but not all) of the groups of interest in physics can be shown to have only type I representations.

Let $L = L^1 \oplus L^2 \oplus \cdots$ where the L^j are irreducible. It is not difficult to show that L is "multiplicity free" in the sense that L^i and L^j are inequivalent whenever $i \neq j$ if and only if $R(L, L)$ is commutative. Now the latter condition makes sense whether L is a direct sum of irreducibles or not and we take it as a definition in the general case. In other words a general representation L will be said to be *multiplicity free* if and only if $R(L, L)$ is commutative. A representation L will be said to be of *type* I if it is equivalent to a direct sum $L^1 \oplus L^2 \oplus \cdots$ where each L^j is multiplicity free. This definition is equivalent to the more usual one which stipulates that $R(L, L)$ should be a ring of type I in the sense of Murray and von Neumann. The problem of determining all type I representations to within equivalence can be reduced to the problem of determining all multiplicity free representations by means of the following theorem.

THEOREM*. *Let L be any representation of type* I. *Then there exists a sequence* $L^\infty, L^1, L^2, \cdots$ *of disjoint multiplicity free representations such that L is equivalent to the representation* $\infty L^\infty \oplus L^1 \oplus 2L^2 \oplus 3L^3 \oplus \cdots$. *The* L^j *are unique to within equivalence. Here nM is an abbreviation for* $M \oplus M \oplus \cdots \oplus M$ *with n terms and some summands may be missing.*

The theory reducing the study of multiplicity free representations to the study of irreducible ones takes its simplest form when G is commutative and will be described in detail in the next section.

4. Multiplicity free representations of commutative groups. In classifying the irreducible representations of the groups in which we shall ultimately be interested we shall need to know something about much more general representations of commutative groups. It can be shown that all representations of commutative groups are of type I. Thus by the theorem of the last section we need only study multiplicity free representations of these groups. We shall show in this section how the classical Hahn-Hellinger theory describing the unitary equivalence classes of self-adjoint operators can be adapted to give a quite complete and satisfying analysis. It is perhaps not without interest to note that, via Stone's theorem, the study of representations of the additive group of the real line is completely equivalent to the study of self-adjoint operators.

It is an easy consequence of Schur's lemma that an irreducible representation L of a commutative group is necessarily one-dimensional. Thus each L_x is of the form $\chi^L(x)I$ where I is the identity operator and χ^L is a continuous function from G to the complex numbers of modulus one such that $\chi^L(xy) = \chi^L(x)\chi^L(y)$ for all x and y in G. Such functions are called *characters* of G. Conversely if χ is any character of G and I is the identity in a one-dimensional Hilbert space, then $x \to \chi(x)I$ is an irreducible representation L^χ of G. Moreover L^{χ_1} and L^{χ_2} are equivalent if and only if χ_1 and χ_2 are identical. Thus determining the equivalence classes of irreducible representations of G in the commutative case is the same as finding the set $\hat{G}$ of all characters. This problem is easily solved in familiar cases. For example, let G be the additive group R_n of all n-tuples of real numbers. Then for each n-tuple $a_1, a_2, \cdots, a_n$ of real numbers we get a character χ by setting $\chi(x_1, x_2, \cdots, x_n) = e^{i(a_1x_1 + \cdots + a_nx_n)}$ and it can be shown that every character of R_n arises in this way from a unique element in R_n.

*For references to the literature in which proofs of this and other theorems will be found see § 12.

As another example consider the multiplicative group T of all complex numbers of absolute value 1. For each integer n we may define a character χ by setting $\chi(Z) = Z^n$ and it can be shown that every character of T is of this form.

In general the set $\hat{G}$ of all characters of G is itself a commutative group. In fact the product of two characters is clearly a character and this operation obviously converts $\hat{G}$ into a group. Moreover it is not hard to show that $\hat{G}$ admits a natural topology under which it is locally compact and separable. $\hat{G}$ is called the group dual to G. The use of the term "dual" is justified by the Pontrjagin duality theorem which we shall now state. Note first that if we hold x fixed in G and let χ vary over the members of $\hat{G}$, then $\chi(x)$ as a function on $\hat{G}$ is a character, i.e., there exists a unique member F_x of $\hat{\hat{G}}$ such that $F_x(\chi) \equiv \chi(x)$. According to Pontrjagin's theorem, every member of $\hat{\hat{G}}$ is of the form F_x for a unique x in G and $x \to F_x$ sets up an isomorphism between G and $\hat{\hat{G}}$ as topological groups. Thus we may identify G with $\hat{\hat{G}}$ and think of G and $\hat{G}$ as being in a mutually reciprocal relationship. Because of this it is customary to write $\chi(x)$ in the more symmetrical form (x, χ). It can happen (for example if G is finite or a vector group) that G and $\hat{G}$ are themselves isomorphic but there is hardly ever a "natural" isomorphism which permits us to identify G and $\hat{G}$.

Now let μ be any measure in the dual $\hat{G}$ of a commutative G which is completely additive, defined on all Borel sets and σ-finite. We define a representation L^μ of G whose space $H(L^\mu)$ is $\mathscr{L}^2(\hat{G}, \mu)$ by setting $(L^\mu_x(f))(\chi) = \chi(x)f(\chi)$. Concerning these representations L^μ it is possible to prove the following theorems.

THEOREM 4.1. *Every L^μ is multiplicity free.*

THEOREM 4.2. *Every multiplicity free representation of G is equivalent to some L^μ.*

THEOREM 4.3. *L^μ and L^ν are equivalent if and only if μ and ν have the same sets of measure zero.*

THEOREM 4.4. *L^μ and L^ν are disjoint if and only if μ and ν are mutually singular; that is, if and only if there exists a Borel subset N of $\hat{G}$ such that $\mu(N) = 0$ and $\nu(\hat{G} - N) = 0$.*

We conclude this section with a few remarks about the analysis of the possible multiplicity free representations of G provided by these theorems. It is easy to see that L^μ is a direct sum of irreducible representations if and only if μ is concentrated in a countable set. When this happens the class of measures having the same null sets as μ is completely described by this countable set which is simply the complement of the unique largest μ null set. In general, however, there is no largest null set and hence no subset of $\hat{G}$ which describes the class of μ or the equivalence class of L^{μ}. It is somewhat illuminating to think of a measure class; i.e., the set of all measures having a given family of null sets as a sort of generalized subset of $\hat{G}$.

5. Semi-direct products. We shall be interested henceforth in groups G which admit a non-trivial factorization of the following sort: there exists a closed normal subgroup N and a second closed subgroup K (not necessarily normal) such that every $x \in G$ may be written uniquely in the form nk where $n \in N$ and $k \in K$. As is easy to see N and K factor G in this way if and only if $NK = G$ and $N \cap K = e$. Let n_1k_1 and n_2k_2 be two elements of G such that $n_i \in N$ and $k_i \in K$. Then $(n_1k_1)(n_2k_2) = n_1k_1n_2k_1^{-1}k_1k_2$ where now $n_1k_1n_2k_1^{-1}$ is in N and k_1k_2 is in K. From this it is easy to see that we can reconstruct G once we are given N, K and the function knk^{-1} from $N \times K$ to N. Of course for each fixed k, $n \to knk^{-1}$ is an automorphism of N and the mapping from K to the group of automorphisms of N is a homomorphism. Thus the function in question may be described as a homomorphism from K to the group of automorphisms of N. Conversely let N and K be any two locally compact groups such that N is commutative (though commutativity of N plays no role in the discussion of this section). Let there be given a homomorphism φ of K into the group of automorphisms of N such that $\varphi(k)(n)$ is continuous on $N \times K$. We may convert the set of all pairs (n, k) in $N \times K$ into a locally compact group by giving it the product topology and defining $(n_1, k_1)(n_2, k_2)$ as $(n_1\varphi(k_1)(n_2), k_1k_2)$. The set of all n, k with k the identity e will then be a closed normal subgroup isomorphic to N and the set of all e, k a "complementary factor" isomorphic to K. When $\varphi(k)$ is the identity for all k we have the ordinary direct product of N and K. In the general case we shall call the result the semi-direct product of N and K defined by φ. Usually we shall suppress the symbol φ and write $\varphi(k)(n)$ as $k(n)$.

There are many interesting examples of semi-direct products. The group of all permutations of three objects is a semi-direct product of the normal subgroup consisting of (a, b, c), (a, c, b) and the identity with the subgroup consisting of (ab) and the identity. In this case φ takes (ab) into the automorphism which interchanges the two non-identity elements in N. A less trivial example is the group of all rigid motions in three space. In this case N is the group of all translations and K may be taken to be the group of all rotations about some fixed point. Here φ takes each rotation into the automorphism of N defined by submitting each translation vector to the rotation in question. In much the same way the inhomogeneous Lorentz group is a semi-direct product of the four-dimensional vector group of all space time translations with the homogeneous Lorentz group.

6. The restriction of an irreducible representation to a commutative normal subgroup. Our principle concern from now on will be to relate the irreducible representations of a semi-direct product $G = N \times_s K$ to those of N and those of certain subgroups of K. The analysis which turns out to be possible leads in many cases to a determination of all the irreducible representations of G. The basic strategy is the following. Given any irreducible representation L of G we may obtain a representation L^0 of N by considering L_x only for x in N. This representation will usually be reducible and we can ask which reducible representations of N are of the form L^0. It turns out to be possible to give a quite complete answer, especially when certain further conditions are satisfied, and we are thus led to an important classification of the irreducible representations of G. The problem remains of studying the class of all irreducible representations of G having a fixed restriction to N and this turns out to be equivalent to determining all irreducible representations of a certain subgroup of K.

Proceeding to details we state the first important fact as a theorem.

THEOREM 6.1. *If L is an irreducible representation of G then the restriction L^0 of L to the closed normal subgroup N is of the form lM where $l = \infty, 0, 1, 2, \cdots$ and M is multiplicity free—in other words L^0 has uniform multiplicity.*

It will be convenient to ignore the l for the time being and concentrate attention on the multiplicity free representation M. As explained in §4 every such M defines and is defined by a uniquely determined measure class in $\hat{N}$. Thus every irreducible L of G defines a unique measure class C_L in $\hat{N}$.

In order to discuss the next question: which measure classes in $\hat{N}$ are of the form C_L for some irreducible L?—we must first make some observations about a certain natural action of K on $\hat{N}$. If α is any automorphism of N (as a topological group) it induces a natural automorphism of $\hat{N}$. Indeed for each $\chi \in \hat{N}$, $n \to \chi(\alpha(n))$ is also a character which we shall denote by $[\chi]\alpha$. Clearly $\chi \to [\chi]\alpha$ is an automorphism of $\hat{N}$. These considerations apply in particular when α is the automorphism $n \to k(n)$ defined by an element k of K. Thus we define $[\chi]k$ for each $\chi \in \hat{N}$ and each $k \in K$. The set of all $[\chi]k$ for fixed χ and k taking on all values in K will be called the *orbit* of χ under K

and will be denoted by $\mathcal{O}_\chi$. It is clear that two orbits $\mathcal{O}_{\chi 1}$ and $\mathcal{O}_{\chi 2}$ either coincide or have nothing in common. Thus $\hat{N}$ is divided up into disjoint sets.

For each measure μ in $\hat{N}$ and each k in K we get a new measure $k(\mu)$ by setting $k(\mu(E)) = \mu(E(k))$. If μ and $k(\mu)$ have the same sets of measure zero for all k we say that μ is *quasi-invariant* under K. In slightly different terminology, μ is quasi-invariant if and only if each $k \in K$ carries each member of the measure class of μ into some other member of this same class. Now just as each k in K carries each μ into some other μ it carries each measure class into some other measure class. From what we have just said it is clear that each $k \in K$ carries a measure class into itself if and only if every member of the class is quasi-invariant under K. Of course, for a given measure class either every member is quasi-invariant or else no member is.

We shall say that a quasi-invariant measure μ in $\hat{N}$ is *ergodic* if no measurable subset is invariant under all $k \in K$ unless it is of measure zero or has a complement of measure zero. Clearly whether or not μ is ergodic depends only upon the class to which it belongs. Thus we may speak of an ergodic invariant measure class. We may now state our second theorem.

THEOREM 6.2. *A measure class in $\hat{N}$ is of the form C_L for some irreducible representation L of G if and only if it is invariant and ergodic under K.*

7. **The ergodic invariant measure class defined by an orbit.** Theorem 6.2 is of course of dubious usefulness unless a method is available for finding the possible ergodic invariant measure classes in $\hat{N}$. In this section we shall describe a method which is applicable whenever the partition of $\hat{N}$ defined by the orbits is "sufficiently smooth." In the general case it can be shown that there is one and only one ergodic invariant measure class concentrated in each orbit. When the appropriate smoothness hypothesis is made it can be shown that every ergodic measure class is concentrated in some orbit. It follows in this case then that we have a natural one-to-one correspondence between orbits and ergodic invariant measure classes.

For each $\chi \in \hat{N}$ let H_χ denote the closed subgroup of K consisting of all k such that $[\chi]k = \chi$. Then $k \to [\chi]k$ defines a one-to-one Borel set preserving map of the space K/H_χ of all right H_χ cosets in K onto the orbit $\mathcal{O}_\chi$ of χ. Let μ be any finite countably additive measure defined on the Borel subsets of K and having the same null sets as Haar measure. For each Borel subset E of K/H_χ let $\tilde{\mu}(E) = \mu(\tilde{E})$ where $\tilde{E}$ is the inverse image of E in K. It is easy to see that $\tilde{\mu}$ is quasi-invariant and ergodic with respect to the natural action (right translation) of K on K/H_χ. Moreover it can be shown that every quasi-invariant measure in K/H_χ is in the same measure class with $\tilde{\mu}$. Thus there is a unique invariant measure class in K/H_χ. Using the one-to-one mapping of K/H_χ on $\mathcal{O}_\chi$ described above we conclude without difficulty that there exists a unique ergodic invariant measure class in $\mathcal{O}_\chi$. Of course an ergodic invariant measure class in $\mathcal{O}_\chi$ is the "same thing" as an ergodic invariant measure class in $\hat{N}$ which is concentrated in $\mathcal{O}_\chi$.

To see that there may exist ergodic invariant measure classes not concentrated in any orbit, consider the following example. Let $\hat{N}$ be the additive group of all complex numbers, let K be the additive group of all integers and let $\varphi(k)$ be the automorphism $z \to ze^{i\alpha\pi k}$ where α is some fixed irrational number. For each $r > 0$ let $\mu_r(E)$ be the (linear) Lebesgue measure of the intersection of E with the circle $|z| = r$. The measure μ_r is invariant under K and is easily shown to be ergodic. On the other hand all of the orbits are countable and the μ_r measure of every countable set is zero. Thus μ_r is not concentrated in any orbit.

Whenever such examples exist the problem of determining all ergodic invariant measure classes is very difficult and is not solved in any case known to the author. Thus it is fortunate that there is a simple, often verified, condition which assures us that there are no such examples.

THEOREM 7.1. *If there exists a Borel subset of $\hat{N}$ which meets each orbit in just one point then every ergodic invariant measure class is concentrated in some orbit.*

The conclusion of Theorem 7.1 is also implied by certain formally weaker assumptions about the orbits in $\hat{N}$. Whenever the *conclusion* holds we shall say that G is a *regular* semi-direct product of N and K.

8. The irreducible representation of G associated with a fixed orbit. When G is a regular semi-direct product we have now reduced the problem of finding all irreducible representations of G to that of finding all irreducible representations L of G for which C_L lies in a fixed orbit $\mathcal{O}$. We shall see that this problem is equivalent to that of finding all irreducible representations of a certain subgroup of K; specifically the subgroup H_χ where χ is any element of $\mathcal{O}$.

In order to make explicit the indicated connection between irreducible representations of H_χ and G it is convenient to make use of the notion of "inducing" a representation from a subgroup to the whole group. Let G be any separable locally compact group and let H be any closed subgroup of G. Let μ be any member of the unique invariant measure class in the coset space G/H. Form the Hilbert space $\mathscr{L}^2(G/H, \mu)$. If μ were actually invariant then each right translation $f(x) \to f(xy)$ would define a unitary operator U_y in this Hilbert space and $y \to U_y$ would be a representation of G. In the general case we can compensate for non-invariance by multiplying by $\rho_y(x)$, the square root of the appropriate Radon-Nikodym derivative relating the translated and untranslated measures. If $(U_y(f))(x) = \rho_x(y)f(xy)$ then U_y is unitary and $y \to U_y$ is a representation. It is easy to show that choosing another μ replaces U by an equivalent representation. This construction is the special case of the one we are interested in in which the inducing representation is the trivial one.

To make the transition to the general case note first that a function on G/H may be regarded as a function on G which is constant on the right H cosets; that is, a function f on G which satisfies the identity:

$$(*) \qquad f(\xi x) = f(x)$$

for all ξ in H and all x in G. Now let μ and ρ_y be as above and suppose that we are given a representation L of H. If we replace complex valued functions by functions having values in the Hilbert space $\mathscr{H}(L)$ we may consider functions in which the identity (*) is replaced by

$$(**) \qquad f(\xi x) = L_\xi(f(x)).$$

If f satisfies (**) then $(f \cdot f)$ satisfies (*) since $(f(\xi x) \cdot f(\xi x)) = (L_\xi(f(x)) \cdot L_\xi f(x)) = (f(x) \cdot f(x))$. Thus if f is also measurable we may consider $(f(x) \cdot f(x))$ as a function on G/H and integrate it with respect to μ. The set of all f for which this integral is finite is a Hilbert space with the integral as the square of the norm. For each f in this Hilbert space and each y in G let $(U_y^L f)(x) = \rho_x(y)f(xy)$ just as above. It is clear that each U_y^L is unitary and it can be shown that $y \to U_y^L$ is a representation of G. To within equivalence this representation, U^L, is independent of μ and depends only upon L. We call it the representation of G *induced* by the representation L of H. In general U^L will not be irreducible even if L is. However, in certain important cases U^L is irreducible and formation of the U^L for suitable subgroups H is one of the chief ways of constructing the irreducible representations of non-commutative non-compact groups.

Returning to our problem, let χ be any element of $\hat{N}$. Form the subgroup NH_χ of G consisting of all products nk with n in N and k in H_χ. For each irreducible representation L of H_χ the correspondence $nk \to \chi(n)L_k$ is easily seen to be a representation of NH_χ. Let us denote it by χL and form the induced representation $U^{\chi L}$ of G.

THEOREM 8.1. *For every irreducible representation L of H_χ the induced representation $U^{\chi L}$ of G is irreducible. The restriction of $U^{\chi L}$ to N has its measure class concentrated in $\mathcal{O}_\chi$ the orbit of χ. Every irreducible representation of G associated with the orbit $\mathcal{O}_\chi$ is of the form $U^{\chi L}$ where L is unique to within equivalence.*

9. Some simple examples. When G is the permutation group on three objects a, b, and c, $\hat{N}$ is a cyclic group of order 3 and the non-identity element of K interchanges the two non-identity elements of $\hat{N}$. Thus there are two orbits in $\hat{N}$. The orbit consisting of the identity above corresponds to the irreducible representations of G which are the identity on N, i.e., to the representations of G obtained by "lifting" the irreducible representations of K. Since K is commutative and of order two, there are two one-dimensional irreducible representations of K. Thus there are two one-dimensional representations of G associated with the identity orbit. Let χ be either one of the elements in the other orbit. Then H_χ consists of the identity above. Thus $U^{\chi I}$ is the only irreducible representation of G associated with the other orbit. It is easily seen to be two-dimensional. These three representations exhaust the irreducible representations of G.

Let G be the group of all rigid motions in the plane considered as a semi-direct product of the translation group N and the group K of rotations about O. The most general χ in $\hat{N}$ is defined by a pair a, b of real numbers—$\chi^{a,b}(x, y) = e^{i(ax+by)}$. The rotation θ takes $\chi^{a,b}$ into $\chi^{c,d}$ where $c = a\cos\theta + b\sin\theta$, $d = -a\sin\theta + b\cos\theta$. Thus $\chi^{a,b}$ and $\chi^{c,d}$ lie in the same orbit if and only if $a^2 + b^2 = c^2 + d^2$. The irreducible representations of G associated with the orbit consisting of 0, 0 are just the countably many one-dimensional irreducible representations of K lifted to G. For each $\chi^{a,b}$ with $a^2 + b^2 > 0$, $H_{\chi^{a,b}}$ reduces to the identity. Hence $U^{\chi^{a,b}}$ is irreducible and is to within equivalence the unique irreducible representation of G associated with the orbit of $\chi^{a,b}$. Thus for each $r > 0$ there is a unique infinite dimensional irreducible representation of G.

10. The inhomogeneous Lorentz group. Next we consider a more complicated example but one of great physical interest—that in which N is the additive group of all quadruples of real numbers x_0, x_1, x_2, x_3, and K is the connected component of the identity in the group of all linear transformations of N onto N which leave fixed the quadratic form $x_0^2 - x_1^2 - x_2^2 - x_3^2$. Each member of K is by definition an automorphism of N so it makes sense to say that φ is the identity homorphism. The resulting semi-direct product $N \times_\varphi K$ is isomorphic to the so-called proper inhomogeneous Lorentz group, the connected component of the identity in the group of all relativistic automorphisms of space time. For each quadruple $p = (p_0, p_1, p_2, p_3)$ of real numbers, let $\chi^{p}(x_0, x_1, x_2, x_3) = e^{i(x_0p_0 - x_1p_1 - x_2p_2 - x_3p_3)}$. Then each χ^p is a member of $\hat{N}$ and every member of $\hat{N}$ may be so obtained. Each $\alpha \in K$ has the form $(x_0, x_1, x_2, x_3) \to \sum_{j=0}^{3} a_{ij}x_j$, and the corresponding automorphism of $\hat{N}$ is defined by $(p_0, p_1, p_2, p_3) \to (\sum_{j=0}^{3} b_{ij}p_j)$ where $b_{ij} = a_{ji}$ for $i, j = 1, 2, 3$, $i = j = 0$, $b_{0,j} = -a_{j,0}$ and $b_{j,0} = -a_{0,j}$ for $j = 1, 2, 3$. It follows that the action of K on $\hat{N}$ is the same as on N—the transformations $p \to [p]\alpha$ are precisely the members of the connected component of the identity in the group of linear transformations leaving fixed the quadratic form $p_0^2 - p_1^2 - p_2^2 - p_3^2$.

It follows that each orbit is contained in one of the hyperboloids $p_0^2 - p_1^2 - p_2^2 - p_3^2 = c$ where c is a real constant. An easy calculation shows that for each $c < 0$ the corresponding hyperboloid is itself an orbit which we may call $\mathcal{O}_c$. On the other hand for each $c > 0$ the hyperboloid is a union of two orbits, one $\mathcal{O}_{+c}$ containing $(\sqrt{c}, 0, 0, 0)$, and one $\mathcal{O}_{c-}$ containing $(-\sqrt{c}, 0, 0, 0)$. For $c = 0$, the hyperboloid degenerates into a cone containing three orbits: one $\mathcal{O}_{00}$ containing 0 alone, one $\mathcal{O}_{0+}$ containing $(1, 1, 0, 0)$ and one $\mathcal{O}_{0-}$ containing $(-1, 1, 0, 0)$. It is obvious that there is a Borel cross section for the orbits so that we have a regular semi-direct product.

It will be convenient to divide the discussion of the representations associated with each orbit into cases.

CASE I. $c > 0$. Both $H_{\chi^{(\sqrt{c},0,0,0)}}$ and $H_{\chi^{(-\sqrt{c},0,0,0)}}$ are isomorphic to the group of all linear transformations in E^3 which have positive determinant and leave $x_1^2 + x_2^2 + x_3^2$ fixed, that is to the rotation group in three dimensions. We have already described the irreducible representations of this group. They are the $2j + 1$ dimensional representations D_j where $j = 0, 1, 2, \cdots$. Thus for each $c > 0$ we get two infinite sequences of irreducible representations of G, $U^{\chi^{(\sqrt{c},0,0,0)}D_j}$ and $U^{\chi^{(-\sqrt{c},0,0,0)}D_j}$. For future reference let us call these $L^{\sqrt{c},j}$ and $L^{-\sqrt{c},j}$ respectively.

CASE II. $\mathcal{O}_{0-}$ and $\mathcal{O}_{0+}$. $H_{\chi^{(1,1,0,0)}}$ and $H_{\chi^{(-1,1,0,0)}}$ are isomorphic to the group of all rigid motions in the plane. This group was discussed in the preceding section and its representations fell into two classes. There was a "continuous" family of infinite dimensional irreducible representations parameterized by the positive real number r, and a "discrete" family of one-dimensional irreducible representations parameterized by an integer n. We have accordingly four classes of irreducible representations of G which we shall denote by $L^{0+,r}$, $L^{0-,r}$, $L^{0+,0,n}$, $L^{0-,0,n}$.

CASE III. $\mathcal{O}_{0,0}$. $H_{\chi^{(0,0,0,0)}} = K$ and the corresponding irreducible representations of G are just those of K lifted to G. The determination of the irreducible representations of K is a difficult problem which has been solved in [3] by Gelfand and Neumark. However, for reasons which will be indicated below the corresponding representations of G are not of physical interest. Accordingly, we shall say no more about them.

CASE IV. $c < 0$. $H_{\chi^{(0,\sqrt{c},0,0)}}$ is isomorphic to the homogeneous Lorentz group in three-dimensional space time. The determination of the irreducible representations of this group is also a difficult problem. It has been solved in [1] by V. Bargmann. Again, however, the corresponding representations of G do not seem to be of physical interest and we shall not discuss them further.

We conclude this section with some brief indications concerning the physical significance of the results described. For plausible reasons which we shall not discuss one expects the representation of the inhomogeneous Lorentz group associated with a relativistic quantum mechanical system to be irreducible when the system is an "elementary particle." Thus the classification of irreducible representations which we have given yields a corresponding classification of elementary particles. Let us examine the physical meaning of the parameters in this classification. As indicated in §1, the energy observable is the infinitesimal generator of the restriction of our representation to the one-parameter group $(x_0, x_1, x_2, x_3) \to (x_0 + t, x_1, x_2, x_3)$. Moreover the x component of momentum observable is the infinitesimal generator of the restriction of our representation to the one-parameter group $(x_0, x_1, x_2, x_3) \to (x_0, x_1 + x, x_2, x_3)$. Similar statements hold for the y and z linear momentum components. Of course we are supposing units chosen so that $\hbar = 1$. It follows that those observables depend only upon the restriction of our representation to N and hence only upon the relevant orbit in $\hat{N}$. The representation of N associated with a given orbit takes the simple form described in §4 and from this it is easy to compute that for an orbit lying in the hyperboloid $p_0^2 - p_1^2 - p_2^2 - p_3^2 = c$, the energy observable P_0 is related to the momentum observables in the x, y, and z direction P_1, P_2, and P_3 by the equation $P_0^2 = c + P_1^2 + P_2^2 + P_3^2$. This is exactly the relationship between energy and momentum for a relativistic particle of rest mass $\sqrt{c}$ whenever $c \geqslant 0$. When $c < 0$ we get a relationship between energy and momentum which does not occur for any physical particle and would imply an imaginary rest mass. This is why Case IV is rejected. In Case III the observables P_0, P_1, P_2, and P_3 are all zero—another physically impossible situation.

The representations $L^{\pm\sqrt{c},j}$ in Case I are described by two parameters. The absolute value of the first has already been correlated with the rest mass of the corresponding particle. Its sign is not physically significant since $L^{\sqrt{c},j}$ and $L^{-\sqrt{c},j}$ are "equivalent" via an anti-unitary transformation. What about j? We shall not give details but simply state that one computes the angular momentum observables and is led to the conclusion that a particle whose representation is $L^{\sqrt{c},j}$ has "spin" j. We get only integral spins because we have ignored the projective representations of G.

In Case II where the particles have rest mass zero the integral parameter can also be interpreted as spin though not in quite so straightforward a way. The representations with the continuous parameter r have a dubious physical significance.

11. Supplementary remarks. Let G be a type I group (i.e., having only type I representations) and let $\tilde{G}$ denote the set of all equivalence classes of irreducible representations of G. It is possible to define a notion of Borel set in $\tilde{G}$ and to extend the considerations of §4 so as to obtain a one-to-one correspondence between measure classes in $\tilde{G}$ and multiplicity free representations of G completely analogous to that obtained in §4 for commutative groups. Taking §3 into account we see that once we know $\tilde{G}$ we know all representations of G, irreducible or not. Since we now know $\tilde{G}$ for many regular semi-direct products G it is of particular interest to know when a regular semi-direct product is of type I.

THEOREM 11.1. *A regular semi-direct product $N \times_{\bullet} K$ is of type* I *if and only if H_χ is of type* I *for all $\chi \in \hat{N}$.*

When G is the inhomogeneous Lorentz group, there are essentially four different possibilities for H_χ as pointed out in §10. In two cases H_χ is a connected semi-simple Lie group and hence of type I by a theorem due to Harish-Chandra. In another H_χ is compact and hence of type I by the Peter-Weyl theorem. In the remaining case H_χ is a regular semi-direct product of two commutative groups and hence is of type I by Theorem 11.1 and the theorem that all commutative groups are of type I. Thus the inhomogeneous Lorentz group is of type I.

What can one say about the irreducible representations of G when

G is an irregular semi-direct product of N and K; for example, when G is the subgroup of all rigid motions in the plane generated by the translations and a rotation through an irrational multiple of π? First of all we can certainly find lots of representations. For each orbit $\mathcal{O}_\chi$ one can form H_χ and the irreducible representations $U^{\chi L}$ just as in the regular case. However two difficult problems remain: (a) that of finding all ergodic invariant measure classes not concentrated in orbits and (b) that of analyzing the set of all irreducible representations associated with a fixed such ergodic invariant measure class. The author is now studying question (b) and has found that one can develop a theory of "virtual subgroups" and their representations to compensate for the lack of existence of an H_χ when the measure class is not concentrated in an orbit.

We remark finally that the theory we have been describing has been extended to the case in which N instead of being commutative is only of type I, K does not necessarily exist, and the representations may be only projective. The problem of finding all representations of the anti-commutation relations may be reduced to a problem in this more general theory but leads to the irregular case.

12. Guide to the literature. The irreducible representations of the inhomogeneous Lorentz group were first found by Wigner in [11]. Their connection with elementary particles and relativistic wave equations is discussed in detail by Wigner and Bargmann in [2]. The general theory of semi-direct products described here was first obtained by the author in [4] as a corollary to a theorem about systems of imprimitivity. A more careful statement with examples will be found in [5]. The more general theory indicated in §11 appears in [7] and [8]. [8] contains a more detailed proof of the main theorem of [4]. A detailed treatment of the material of these chapters is included in the lecture notes [6] of a course given by the author at the University of Chicago in 1955. These notes are perhaps easier to read than either [4] or [8]. In all treatments the hypothesis of type I-ness is supplemented by one requiring that G be "smooth." This hypothesis has been shown to be unnecessary by recent work of J. Glimm.

References

1. V. Bargmann, *Irreducible unitary representations of the Lorentz group*, Ann. of Math. (2) **48** (1947), 568–640.

2. V. Bargmann and E. Wigner, *Group theoretical discussion of relativistic wave equations*, Proc. Amer. Acad. Arts Sci. **34** (1948), 211–223.

3. I. M. Gelfand and M. A. Neumark, *Unitary representations of the Lorentz group*, Izv. Akad. Nauk SSSR Ser. Mat. **11** (1947), 411–504. (Russian.)

4. G. W. Mackey, *Imprimitivity for representations of locally compact groups*. I, Proc. Nat. Acad. Sci. U.S.A. **35** (1949), 537–545.

5. ———, *Induced representations of locally compact groups*. I, Ann. of Math. (2) **55** (1952), 101–139.

6. ———, *Theory of group representations*, Lecture Notes, Univ. of Chicago, Summer 1955.

7. ———, *Borel structure in groups and their duals*, Trans. Amer. Math. Soc. **85** (1957), 134–165.

8. ———, *Unitary representations of group extensions*. I, Acta Math. **99** (1958), 265–311.

9. F. Mautner, *Unitary representations of locally compact groups*, Ann. of Math. (2) **51** (1950), 1–25.

10. J. von Neumann, *On rings of operators. Reduction theory*, Ann. of Math. (2) **50** (1949), 401–485.

11. E. Wigner, *On unitary representations of the inhomogeneous Lorentz group*, Ann. of Math. (2) **40** (1939), 149–204.

Index